

**EIN DIGITALES SPRACHINTERPOLATIONSVERFAHREN
MIT PRÄDIKTIONSGESTEUERTER WORTAUFTeilUNG**

Heinz Gerhäuser

Erlangen, 1980



Digitized by the Internet Archive
in 2026

EIN DIGITALES SPRACHINTERPOLATIONSVERFAHREN MIT PRÄDIKTIONSGESTEUERTER WORTAUFTeilUNG

Der Technischen Fakultät
der Universität Erlangen-Nürnberg
zur Erlangung des Grades
DOKTOR-INGENIEUR

vorgelegt von
Heinz Gerhäuser

Erlangen, 1980



Als Dissertation genehmigt
von der Technischen Fakultät
Der Universität Erlangen-Nürnberg

Tag der Einreichung: 6. Mai 1980
Tag der Promotion: 25. Juni 1980

Dekan:	Prof. Dr.-Ing. H. Niemann
Berichterstatter:	Prof. Dr.-Ing. D. Seitzer
	Prof. Dr.-Ing. W. Schüßler

Inhaltsverzeichnis

1.	Einleitung	1
2.	Stand der Technik	3
2.1	Datenreduktion durch Quellencodierung	3
2.2	Verfahren zur Quellencodierung	7
2.3	Datenreduktion durch Ausnützen der Sprachpausen	13
2.3.1	Prinzip der Sprachinterpolation	13
2.3.2	Sprachinterpolationsverfahren	16
2.4	Neue Beiträge	20
3.	Einführung in das PWA-Verfahren	22
3.1	Grundlegende Idee	22
3.2	Der unregelmäßige Datenfluß bei Sprachsignalen	24
3.3	Festlegung von Qualitätskriterien	31
4.	Die Anwendung des PWA-Prinzips	34
4.1	PWA1- Sprachinterpolation mit prädiktions- gesteuerter Wortaufteilung ohne Sprachdetektor	34
4.1.1	Klassifizierung des Sprachsignals	35
4.1.2	Bestimmung der bedingten Häufigkeitsverteilungen für PCM- und DPCM-Werte	42

4.1.3	Berechnung von optimalen Quantisierungskennlinien	43
4.1.4	Codierung der Einzelquelle	47
4.1.5	Extrapolation des Übertragungsbedarfs bei Sprachinterpolation	50
4.2	PWA2- Ein verbessertes PWA-Verfahren	52
4.2.1	Einfluß des segmentweisen Pausenausschlusses auf die Aktivität	53
4.2.2	Unterscheidung zwischen Normalbetrieb und segmentweiser PCM-Übertragung	55
4.2.3	Codierung mit Pegelregelung	58
4.2.4	Signalklassifizierung bei Codierung mit Pegelregelung	65
4.2.5	Codierung der Einzelquelle	67
4.2.6	PWA2- Simulation mit Scharmittelung über eine endliche Quellenzahl	72
4.3	Zusammenfassung des Abschnitts 4	75
4.3.1	Das PWA1-Verfahren	75
4.3.2	Das PWA2-Verfahren	77
4.4	Maßnahmen bei Fehlern auf dem Übertragungskanal	79
5.	Computersimulation	81
5.1	Sprachsignal- Ein/Ausgabe am Institutsrechner (HP 2100 A)	81
5.2	Verwendete Sprachproben	84

5.3	Programmierung der Systemsimulation	85
6.	Ein schneller Signalprozessor für die Echtzeit- realisierung von Sprachverarbeitungssystemen	86
6.1	Architektur des Signalprozessors	87
6.2	Leistungsfähigkeit des Signalprozessors	90
6.3	Programmierung des Signalprozessors	92
6.4	Realisierung von Teilaufgaben im PWA2-System	93
7.	Zusammenfassung	94
8.	Literatur	97
9.	Anhang	108
10.	Verzeichnis der Abkürzungen und Formelzeichen	112

EIN DIGITALES SPRACHINTERPOLATIONSVERFAHREN MIT PRÄDIKTIONSGESTEUERTER WORTAUFTeilUNG

1. Einleitung

Die stürmische Entwicklung der Nachrichtentechnik und Elektronik in den letzten Jahren hat eine Vielfalt neuer Kommunikationsmöglichkeiten mit sich gebracht. Als Beispiele seien hier das Bildfernsprechen, das Fernkopieren, die Datenübertragung oder automatische Auskunftssysteme angesprochen.

Noch befinden sich viele dieser Möglichkeiten in der Erprobungsphase und sind nur einem begrenzten Benutzerkreis zugänglich. Der Zeitpunkt ist jedoch abzusehen, wo eine breite Anwendung wirtschaftlich wird.

Dieser Entwicklung kommt noch entgegen, daß die Digitalisierung des Fernsprechnetzes durch die Einführung der Pulsmodulation (PCM) auch geeignete Übertragungswege entstehen läßt. Es ist deshalb mit einem sprunghaftigen Anwachsen der zu übertragenden Informationsmengen zu rechnen.

Nicht zuletzt zur Vermeidung von Engpässen wird eine wirtschaftliche Informationsübertragung notwendig und der Einsatz von datenreduzierenden Übertragungsverfahren interessant

Die digitale Übertragung von Sprachsignalen wird in der Zukunft einerseits wegen der großen Zahl bereits bestehender Fernsprechverbindungen, andererseits wegen zunehmender Möglichkeiten der "Mensch Maschinen-Kommunikation" via Sprache, eine dominierende Stellung einnehmen. Damit gewinnen datenreduzierende Übertragungsverfahren für digitale Sprachsignale an Bedeutung.

Neben der Möglichkeit, die Datenrate einer Signalquelle durch Redundanz- und Irrelevanzreduktion zu vermindern, bietet sich bei Sprachsignalen auch noch an, die Sprachpausen mit reduzierter Wortlänge zu übertragen oder sie vollständig von der Übertragung auszuschließen. Bei dem entstehenden unregelmäßigen Datenfluß liegt die mittlere Datenrate gegenüber den Maximalwerten um den Faktor 2 bis 3 niedriger. Übertragungsverfahren, die diese Eigenschaft für eine Datenreduktion ausnützen, sind unter dem Begriff Sprachinterpolationsverfahren (speech interpolation) bekannt geworden.

Das im folgenden beschriebene Sprachinterpolationsverfahren benutzt als Referenz das PCM-Übertragungssystem PCM30/32 (8 bit pro Abtastwert, 8 kHz Abtastfrequenz, nichtlineare Kompan-dierung nach der 13-Segmentkennlinie). Bei einer objektiven Qualität, die die gleiche Toleranzvorschrift erfüllt (Signal/Rauschverhältnis (SNR) über dem Signalpegel) wird die Datenrate von 64 kbit/s bei PCM auf < 24 kbit/s beim Sprachinterpolationsverfahren mit prädiktionsgesteuerter Wortaufteilung (PWA) vermindert.

2. Stand der Technik

In diesem Abschnitt wird ein Überblick über die wichtigsten Datenreduktionsverfahren bei der digitalen Sprachsignalübertragung gegeben.

Nach allgemeinen Überlegungen zur Quellencodierung in 2.1, werden in 2.2 Quellencodierverfahren für Einzelquellen ohne Pausenausnutzung betrachtet. Die Prinzipien der Datenreduktion durch Ausnutzung der Sprachpausen und die darauf aufbauenden Sprachinterpolationsverfahren werden im Abschnitt 2.3 behandelt. Schließlich wird in 2.4 der eigene Beitrag skizziert.

2.1 Datenreduktion durch Quellencodierung

Ein Nachrichtenübertragungssystem läßt sich allgemein durch das in Bild 1 gezeigte Modell /2.1-1/ darstellen.

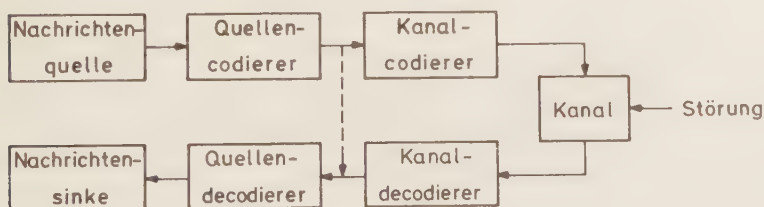


Bild 1: Modell eines Nachrichtenübertragungssystems

In dieser Arbeit wird ein ungestörter Übertragungskanal zugrunde gelegt, so daß Kanalcodierung und -decodierung außer acht gelassen werden können. Das Übertragungssystem vereinfacht sich dadurch. Quellencodierer und Quellendecodierer sind entsprechend dem gestrichelten Pfad in Bild 1 zu verbinden. Die Nachrichtenquelle erzeugt ein digitales Ausgangssignal,

das durch Abtastung eines bandbegrenzten analogen Signals mit der Grenzfrequenz f und einer Quantisierung in N Stufen entstanden sei. Bei Darstellung der N Stufen mit einem Binärkode sind w bit erforderlich. Wenn möglich, ist N so zu wählen, daß die Beziehung gilt: $2^w = N$

Es wird vorausgesetzt, daß die Abtastfrequenz f_A die Bedingung des Abtasttheorems erfüllt:

$$f_A \geq 2f_g$$

Von der Quelle wird eine Nachricht H_0 (Entscheidungsgehalt) abgegeben.

$$H_0 = \lg N$$

H_0 kann als Summe der Entropie H (Informationsgehalt) und der Redundanz R geschrieben werden /2.1-2/.

$$H_0 = H + R$$

Ein Teil der von der Quelle abgegebenen Nachricht ist irrelevant. Darunter versteht man den Nachrichtenanteil, dessen Unterdrückung vom Empfänger nicht bemerkt wird oder toleriert werden kann. Nach Shouten /2.1-3/ läßt sich jede Nachricht in vier Bereiche zerlegen. Dies kann in der sogenannten Nachrichtenebene anschaulich dargestellt werden (Bild 2).



Bild 2: Nachrichtenebene nach Shouten

Ziel der Quellencodierung ist es, zum Empfänger nur den relevanten und nichtredundanten Nachrichtenanteil zu übertragen.

Für die Redundanz in einer Nachricht sind im wesentlichen zwei Ursachen denkbar:

-Redundanz durch statistische Abhängigkeit der Quellensymbole.

-Codedredundanz durch ungleiche Auftretswahrscheinlichkeit der Symbole.

Der erste Fall bezieht sich auf Quellen mit Gedächtnis. Geht man davon aus, daß es sich um eine lineare statistische Abhängigkeit handelt, so kann durch lineare Prädiktion /2.1-4/ eine Dekorrelation erreicht werden. Damit ist eine Überführung in eine gedächtnislose Quelle möglich. Die spektrale Leistungsdichte eines Signals nach vollständiger Dekorrelation ist konstant (weißes Spektrum). Es leuchtet ein, daß die durch Korrelation bedingte Redundanz um so größer ist, je mehr das Spektrum von einem konstanten Verlauf abweicht.

Im zweiten Fall kann gezeigt werden /2.1-5/, daß die Entropie H einer Quelle

$$H = - \sum_{i=1}^N p(u_i) \cdot \lg p(u_i)$$

genau dann maximal wird, wenn die Auftretswahrscheinlichkeiten $p(u_i)$ für die Quellensymbole u_i alle gleich groß sind:

$$p(u_i) = p(u_j) = \frac{1}{N} \text{ für alle } i, j$$

Die Codedredundanz ergibt sich zu

$$R = H_0 - H$$

Mit $H_0 = \lg N$ erhält man

$$R = \lg N + \sum_{i=1}^N p(u_i) \cdot \lg p(u_i)$$

und bei gleicher Symbolwahrscheinlichkeit

$$R = \lg N - \sum_{i=1}^N \frac{1}{N} \cdot \lg N = 0$$

Auch hier kann etwas vereinfacht formuliert werden: Je mehr die Symbolwahrscheinlichkeiten von einer Gleichverteilung abweichen, um so größer ist die Codedredundanz.

Eine optimale Codierung von Symbolen mit unterschiedlicher Auftretswahrscheinlichkeit kann durch Verwendung unterschiedlich langer Codeworte erreicht werden. Bei dieser Entropiecodierung werden häufigere Symbole durch kurze, seltenere Symbole durch längere Codeworte dargestellt. Im Mittel wird die Datenrate minimal. Huffman hat für diesen Code eine Konstruk-

tionsvorschrift angegeben /2.1-6/.

Bei der Irrelevanzreduktion steht die Frage im Vordergrund, ab wann eine genauere Darstellung der Nachricht vom Empfänger nicht mehr ausgenutzt werden kann. Die Irrelevanzreduktion beginnt streng genommen bei der Bandbegrenzung des Sprachsignals vor der Abtastung. Auch die Digitalisierung des Analogsignals ist eine Irrelevanzreduktion.

Die subjektive Beurteilung des Empfängers ist von großer Bedeutung. Man kann z.B. Verdeckungseffekte des menschlichen Gehörs ausnutzen, um entsprechend gefilterte Störsignale weniger in Erscheinung treten zu lassen.

Als anderes Beispiel sei die Abhängigkeit der Hörschwelle eines Störsignals vom Pegel des Nutzsignals angesprochen. Bei der subjektiven Beurteilung von codierten Sprachsignalen ist allerdings zu beachten, daß zwischen Qualität und Verständlichkeit unterschieden werden muß. Nach /2.1-8/ ist nicht automatisch mit einer guten Qualität eine hohe Verständlichkeit und umgekehrt verbunden.

Im Gegensatz zur Redundanzreduktion ist die Irrelevanzreduktion ein irreversibler Vorgang. Es entsteht ein Fehler.

Zwischen dem bei einer Codierung auftretenden mittleren quadratischen Fehler d und der aufgewendeten Datenrate R_D besteht ein Zusammenhang. Die "Rate Distortion Function" /2.1-7/ gibt für ein Signal mit bekannten statistischen Eigenschaften bei einem vorgegebenen Fehler einen unteren Grenzwert für die dazu notwendige Datenrate R_D an (Bild 3).

R_D ist als Produkt aus der Abtastfrequenz f_A und der Bitzahl pro Abtastwert w definiert.

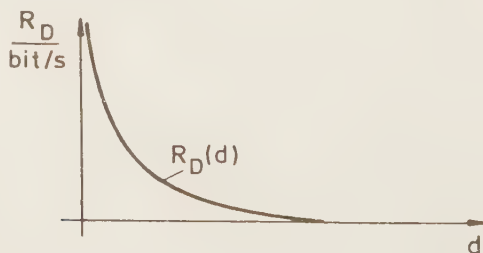


Bild 3: Typische Rate Distortion Function

Allerdings ist die Aussage der Rate Distortion Function mehr von theoretischer Bedeutung, da sie keine Codiervorschrift liefert, mit der diese minimale Datenrate erreicht wird.

2.2 Verfahren zur Quellencodierung

Eine große Zahl von Quellencodierverfahren läßt sich zu einer Gruppe zusammenfassen. Gemeint sind Verfahren zur Signalformcodierung (Waveform Coding). Ziel dabei ist, die Signalform so genau wie möglich zu übertragen. Für einen Teil dieser Verfahren ist es charakteristisch, daß der Störabstand SNR (Signal to Noise Ratio) als objektives Qualitätskriterium eine gute Relation zur subjektiven Beurteilung durch Hörvergleich darstellt. Dies gilt vor allem, wenn das Störgeräusch weißem Rauschen ähnlich ist.

Die lineare Pulsmodulation (PCM) kann als einfachste Signalformcodierung betrachtet werden /2.2-1/. Da bei diesem Verfahren Signaleigenschaften nicht berücksichtigt werden, kann eine Datenreduktion nur im Rahmen einer Irrelevanzreduktion erfolgen. Die lineare PCM hat für die Übertragung jedoch kaum Bedeutung. Um einen Ruhegeräuschabstand nach CCITT von > 62 dB (gemessen bei Vollaussteuerung) in einem Telefonübertragungssystem einhalten zu können, benötigt man 2048 Quantisierungsstufen /2.2-2/. Die dafür erforderlichen 11 bit führen bei einer Abtastfrequenz von 8 kHz auf eine Datenrate von 88 kbit/s. Von größerer Bedeutung ist, daß bei den meisten Quellcodierverfahren eine Analog/Digitalumsetzung entsprechend einer linearen PCM mit 12 bit vor der Verarbeitung durchgeführt wird. Der dabei auftretende Quantisierungsfehler kann dann gegenüber den Fehlern des Quellcodierverfahrens vernachlässigt werden.

Ein sehr wichtiges Verfahren stellt die Pulsmodulation mit logarithmischer Kompondierung dar /2.2-3/. Hier wird eine Datenreduktion durch eine an das menschliche Ohr angepaßte Irrelevanzreduktion vorgenommen. Dabei wird ausgenützt, daß die Hörschwelle für ein Störgeräusch um so größer ist, je höher der Nutzsignalpegel ist. Daraus ergibt sich die Forderung nach einem konstanten Störabstand über einem möglichst großen Dynamikbereich des Signals. Dies läßt sich mit einer logarithmischen Quantisierungskennlinie erreichen, die abschnittsweise durch 13 lineare Kennlinien approximiert wird. Die Steigung von Kennlinie zu Kennlinie nimmt zu größeren Aussteuerungen hin jeweils um den Faktor 2 ab. Dadurch erhält man über einen Aussteuerbereich von ca. 30 dB einen konstanten Störabstand von 38 dB. Die Wortlänge bei dieser Codierung ist 8 bit, sodaß bei einer Abtastfrequenz von 8 kHz die Datenrate 64 kbit/s beträgt.

Unter der Bezeichnung PCM 30/32 wird dieses Verfahren im Bereich der Deutschen Bundespost /2.2-4/ und voraussichtlich auch in den meisten europäischen Ländern für die digitale Übertragung von Telefongesprächen eingesetzt.

Durch die geplante Einführung im großen Umfang stellt die PCM mit logarithmischer Kompondierung eine Art Standard dar, an dem man konkurrierende Quellencodierverfahren messen sollte.

Bei der Differenzpulsmodulation (DPCM) steht eine Redundanzreduktion durch lineare Prädiktion zur Datenreduktion im Vordergrund /2.2-5/. Dabei wird aus der Signalvergangenheit ein Schätzwert $x(k)$ berechnet. Der Prädiktionsgrad ist durch die Zahl der aus der Vergangenheit berücksichtigten Werte und deren Gewichtung durch die Prädiktionskoeffizienten gegeben. Die Differenz zwischen dem aktuellen, zur Übertragung anstehenden Abtastwert $u(k)$ und dem Schätzwert $x(k)$ ergibt den sogenannten Prädiktionsfehler $d(k)$, der schließlich nach einer Quantisierung als $d_1(k)$ zum Empfänger übertragen wird (Bild 4).

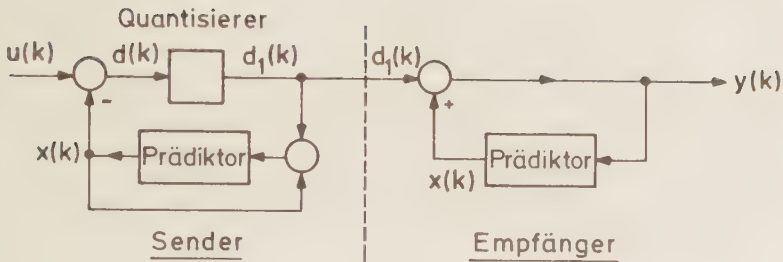


Bild 4: Blockschaltbild eines DPCM-Systems

Auf der Empfangsseite kann der Schätzwert $x(k)$ in gleicher Weise wie auf der Sende- (aus bereits übertragenen Werten) berechnet werden. Durch Addition des übertragenen Prädiktionsfehlers $d(k)$ zum Schätzwert $x(k)$ erhält man das Ausgangssignal $y(k)$. Unter der Voraussetzung eines störungsfreien Kanals weicht das Ausgangssignal $y(k)$ nur um den Quantisierungsfehler $\Delta d(k) = d_1(k) - d(k)$ vom Originalsignal $u(k)$ ab. Die Varianz des Prädiktionsfehlers $d(k)$ wird erst bei vollständiger Dekorrelation minimal. Im allgemeinsten Fall ist dies nur mit unendlich großem Prädiktorgrad möglich. In /2.2-6/ wird jedoch gezeigt, daß bei Sprachsignalen bereits mit Prädiktorgraden von 2 bis 3 der wesentlichste Anteil an Dekorrelation erreicht werden kann.

Als ein Spezialfall der DPCM kann die Deltamodulation (DM) /2.2-7/ betrachtet werden. Vom Prädiktionsfehler wird dann nur noch das Vorzeichen übertragen (1 bit). Allerdings ist dabei eine sehr viel höhere Abtastfrequenz (40...120 kHz) nötig. Obwohl die Übertragungsqualität wegen der für die Deltamodulation typischen Fehler, wie Steigungsrauschen und Ruhegeräusch, nicht mit der 13-Segment PCM konkurrieren kann, ist ein Einsatz in Sondernetzen wahrscheinlich. Der Grund liegt in einer hohen Unempfindlichkeit gegenüber Kanalstörungen. Weiterhin entfällt

hier der sonst für die Wortsynchronisation notwendige Aufwand.

Ein sehr interessantes Sprachcodierverfahren stellt der geschaltete Quantisierer (Switched Quantizer) /2.2-8/ dar. Hauptsächliches Anliegen dieses Verfahrens ist, neben der Dekorrelation durch lineare Prädiktion auch statistische Abhängigkeiten höherer Ordnung zu erfassen. Dabei wird unter der Randbedingung einer konstanten Datenrate der Quantisierer abhängig vom Quellenzustand auf verschiedene Quantisierungskennlinien umgeschaltet.

Die bisher besprochenen dekorrelierenden Methoden benutzen feste Prädiktionskoeffizienten. Bei der Auswahl dieser Koeffizienten muß ein Kompromiß geschlossen werden. Es hat sich gezeigt, daß die Optimierung auf eine Sprachprobe bei anderen Sprachproben zu ungünstigen Ergebnissen führen kann /2.2-9/. Ein Ausweg aus diesem Problem stellen die adaptiven DPCM-Verfahren dar. Allerdings bedingt die kontinuierliche Optimierung der Prädiktionskoeffizienten erheblich komplexere Systeme. Weiterhin muß dem Empfänger der Zustand des Senders mitgeteilt werden. Die dazu notwendige Zusatzinformation lohnt sich jedoch nur, wenn sie nicht für jeden Abtastwert, sondern jeweils für Blöcke von Abtastwerten gültig ist.

Die Adaption kann auch aus der Signalvergangenheit, die dem Empfänger ebenfalls bekannt ist, abgeleitet werden. Die Zusatzinformation zum Empfänger entfällt dadurch. Die momentanen Signaleigenschaften werden dann jedoch nicht gemessen, sondern nur geschätzt. Eine optimale Anpassung ist nicht mehr möglich.

Neben den im Zeitbereich arbeitenden Codierverfahren (Time Domain Coder) sind eine Reihe von Transformationscodiermethoden bekannt /2.2-10/. Es wird nicht das Signal selbst, sondern eine lineare Abbildung codiert. Ausführliche Untersuchungen /2.2-11/ haben gezeigt, daß mit der Walsh-Hadamard-Transformation (WHT) ein Gewinn gegenüber PCM von 3 dB, über die Diskrete Slant-Transformation (DST) ca. 4 dB, die Diskrete Fouriertransformation (DFT) 8 dB, die Diskrete Cosinus-Transformation und die Karhunen-Loeve-Transformation (KLT) ein Ge-

winn von 9 dB erreicht werden kann.

Die KLT liefert den größtmöglichen Gewinn /2.2-12/, benötigt aber einen sehr hohen Rechenaufwand. Günstiger ist die DCT, die bei fast gleichem Gewinn erheblich einfacher zu berechnen ist.

Die Adaptive Transformationscodierung (ATC) /2.2-13/ kombiniert Dekorrelation und Irrelevanzreduktion in eleganter Weise, indem nach einer DCT die verfügbare Datenrate auf die "wichtigsten" Spektralanteile konzentriert wird. Man erhält dadurch auch bei niedrigen Datenraten (< 20 kbit/s) noch eine gute Übertragungsqualität.

Im Gegensatz zu den bisher besprochenen Signalformcodierverfahren steht bei der Teilbandcodierung nicht mehr die genaue Rekonstruktion im Vordergrund, sondern nur noch die subjektive Qualität.

Das Sprachsignalspektrum wird durch Bandpaßfilterung in einzelne Teilbänder aufgespaltet. Entsprechend der einzelnen Bandbreiten kann jeder Teilbereich mit kleinerer Abtastrate übertragen werden. Die Quantisierung in den einzelnen Bändern wird nach subjektiven Kriterien mit unterschiedlichen Wortlängen vorgenommen.

Die Angabe eines Störabstandes ist hier, wie auch bei den Vocoderverfahren, nicht sinnvoll.

Bei sehr niedrigen Datenraten (< 2 kbit/s) können nur noch Vocoderverfahren eingesetzt werden /2.2-15/. Beim Vocoder wird ein Modell für den Prozeß der Spracherzeugung im menschlichen Vokaltrakt zugrunde gelegt. Im Gegensatz zu den Signalformcodierverfahren wird hier nicht die Dynamik der Sprachsignalform, sondern die Dynamik des Spracherzeugungsprozesses übertragen.

Nach einer Signalanalyse auf der Senderseite wird eine parametrische Darstellung des Sprachsignals zum Empfänger übertragen. Dort wird daraus eine Signalsynthese durchgeführt. Sprecherspezifische Eigenschaften des Sprachsignals gehen dabei verloren.

In Tabelle 1 ist ein Überblick über Sprachcodierverfahren mit

typischen Kennwerten und Literaturangaben zusammengestellt. Der aus /2.2-16/ entnommene relative Realisierungsaufwand bezieht sich auf Logik-Gatterfunktionen.

Verfahren	Datenrate in kbit/s	SNR in dB	Qualität	Realisierungsaufwand	Literatur
13-Seg.PCM	64	38	Telefonqualität	1	/2.2-3/
ADM	40	19		1	/2.2-7/
ADPCM	32	30	(Toll-Quality)	1	"
SUB-BAND	24	+		5	/2.2-16/
ATC	16	23		50	/2.2-13/
CV, LPC	2.4	+	"Synthetischer" Klang	100	/2.2-16/
FORMANT	0.5	+		500	"

Tabelle 1: Überblick über Sprachcodierverfahren

(+ bedeutet: Störabstandsangabe nicht sinnvoll)

13-Seg.-PCM: Pulscodemodulation mit nichtlinearer Kompondierung nach der 13-Seg.-Kennl.

ADM: Adaptive Deltamodulation

ADPCM: Adaptive Differenz-Pulscodemodulation

SUB-BAND: Teilbandcodierung

ATC: Adaptive Transformationscodierung

CV: Kanalvocoder

LPC: Linear Predictive Coefficient Vocoder

FORMANT: Formantvocoder

2.3 Datenreduktion durch Ausnützung der Sprachpausen

2.3.1 Prinzip der Sprachinterpolation

Der Begriff Sprachinterpolation wurde in Anlehnung an die englische Bezeichnung "speech interpolation" geprägt. Man versteht darunter die gemeinsame Benutzung von Übertragungskanälen durch mehr Teilnehmer als Leitungen vorhanden sind. Bei einem normalen Telefongespräch wechseln sich die Teilnehmer mit Sprechen und Zuhören ab. Durch zusätzliche Pausen zwischen Worten und Sätzen sinkt die mittlere Belegung eines Übertragungsweges auf ca. 40% ab. Als Maß für die mittlere Belegung wird die "Aktivität einer Quelle" a eingeführt. Typische Werte für a sind:

$a = 80 \dots 90\%$ bei durchgehend gesprochenem Text

$a = 30 \dots 45\%$ bei Telefongesprächen

Das Prinzip von Sprachinterpolationsverfahren beruht darauf, den zeitlich schwankenden Datenfluß durch Mittelung zu glätten und damit die Maximalwerte, die den erforderlichen Übertragungsbedarf bestimmen, zu reduzieren.

Eine zeitliche Mittelung mit Pufferspeichern kommt wegen der dabei notwendigen Verzögerung bis zu mehreren Sekunden nicht in Frage. Außerdem wäre der damit verbundene Aufwand für Datenspeicher erheblich. Gerade bei Telefonverbindungen bietet sich jedoch eine andere Möglichkeit, die Scharmittelung, an. Dabei wird zu jedem Übertragungszeitpunkt über mehrere Teilnehmer gemittelt. Wie in Bild 5 dargestellt ist, verringert sich dadurch die Schwankung der Systembelastung (2.3-1). Unter der Systembelastung B soll die Zahl gleichzeitig aktiver Quellen in einem System mit n Quellen verstanden werden. Daneben soll noch die normierte Systembelastung B_n eingeführt werden, die auf die Gesamtquellenzahl bezogen ist.

$$B_n = B / n$$

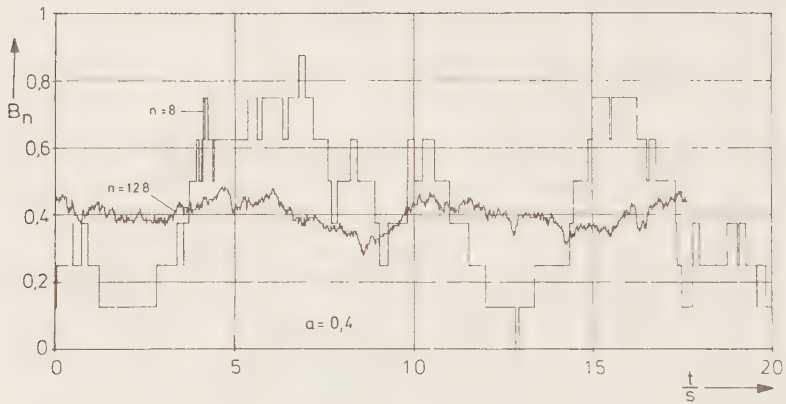


Bild 5: Zeitlicher Verlauf der auf die Quellenzahl n normierten Systembelastung B

Wenn nur eine Quelle vorhanden ist, dann kann B_n nur die Werte 0 oder 100% annehmen. Bereits bei 8 Quellen werden diese Extremwerte nur noch sehr selten erreicht, die Variation von B ist geringer geworden. Besonders eindrucksvoll wird die Glättung, wenn man zu großen Werten für n übergeht. Bei $n = 128$ ist die Schwankung auf den Bereich zwischen ca. 30% und 50% beschränkt. Der Mittelwert von 40% stimmt mit der Aktivität der verwendeten Quellen überein. Bei dieser Betrachtungsweise, wo nur zwischen aktiven und nicht aktiven Teilnehmern unterschieden wird, kann die Systembelastung durch eine Binominalverteilung beschrieben werden. Die Wahrscheinlichkeit p , daß in einem System mit n Quellen, die alle die Aktivität a aufweisen, eine bestimmte Systembelastung B auftritt, ist:

$$p(B) = \binom{n}{B} \cdot a^B \cdot (1-a)^{n-B}$$

Die Verteilungsfunktion $F(x) = P\{x \leq B\}$

ist für verschiedene Quellenzahlen n in Bild 6 dargestellt. Aus Vergleichsgründen wurde auf der Abszisse nicht die Systembelastung B sondern die normierte Systembelastung B_n aufgetragen.

Die Abnahme der Streuung von B_n um den theoretischen Grenzwert von 0.4 für wachsende Quellenzahl n ist deutlich zu erkennen.

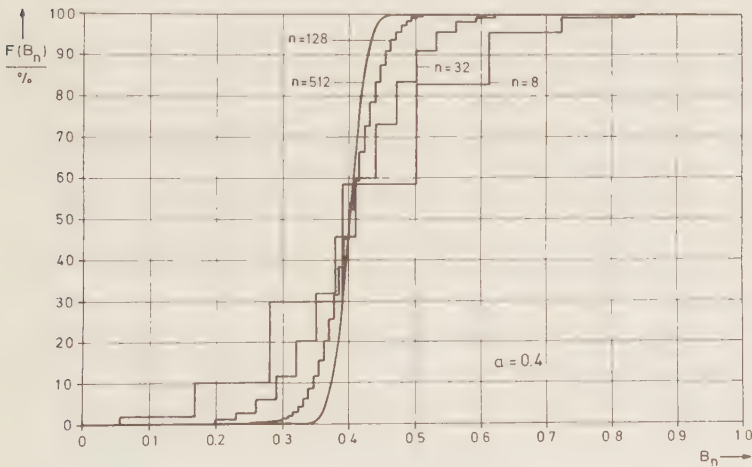


Bild 6: Summenhäufigkeit der Systembelastung

In Sprachinterpolationssystemen nutzt man aus, daß eine bestimmte Systembelastung B_n ab einer genügend großen Quellenzahl n mit großer Wahrscheinlichkeit (z. B. $P = 90\%$ oder $P = 99\%$) nicht mehr überschritten wird. Dieser Zusammenhang ist in Bild 7 dargestellt. In einem System mit 128 Quellen kann die Übertragskapazität gegenüber einem System mit nur einer Quelle um den Faktor 2 reduziert werden, ohne daß mehr als 1% Überlastungen auftreten. Läßt man 10% Überlastungen zu, so ist eine Halbierung der Übertragskapazität bereits bei $n = 32$ möglich.

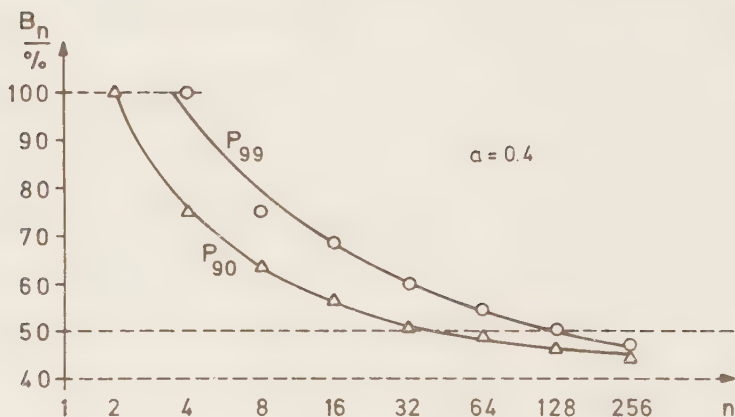


Bild 7: Belastungsgrenzkurven für 10 % und 1 % Überschreitung

Ein wesentliches Unterscheidungsmerkmal der verschiedenen Sprachinterpolationsverfahren ist das Systemverhalten bei Überlastung.

2.3.2 Sprachinterpolationsverfahren

Bereits 1946 wurden die ersten Überlegungen angestellt, wie man vorhandene Übertragungsleitungen, unter Berücksichtigung der Sprachpausen, mehrfach ausnützen kann. Daraus resultierte die Entwicklung des ältesten Sprachinterpolationsverfahrens TASI (Time Assignment Speech Interpolation). Das erste TASI - System konnte 1960 im Transatlantiktelefonieverkehr zwischen London und New York den Betrieb aufnehmen /2.3-2/ Damit wurde es möglich, über 36 Telefonkanäle bis zu 72 Gespräche zu übertragen. Bei diesem in Analogtechnik realisierten System wird durch einen Sprachdetektor festgestellt, welche der 72 Teilnehmer momentan sprechen. Dies muß der Empfangseite durch zu-

sätzliche Information signalisiert werden. Die damit verbundenen Verzögerungen führen zu einem Abschneiden der Sprachanfänge um etwa 20ms. Sprechen mehr als 36 Teilnehmer gleichzeitig, so entstehen für das TASI-System typische Überlastungsstörungen (speech clipping, freeze out). Der Grund dafür ist, daß nach einer Pause ein Teilnehmer nicht sofort eine freie Leitung vorfindet. Bis zur Pause eines anderen Teilnehmers wird auch hier Sprache abgeschnitten. Tritt dies nur sehr selten (0,1%) auf, dann wird diese Störung nicht wahrgenommen. Sie tritt erst ab etwa 2% störend in Erscheinung. Vom klassischen TASI-Verfahren wurde 1970 eine, ebenfalls analog arbeitende, verbesserte Version (TASI-B) /2.3-3/ veröffentlicht, die vor allem durch eine größere Quellenzahl eine bessere Ausnutzung der Signalstatistik ermöglicht.

Das erste digitale TASI-System /2.3-4/ lehnt sich noch eng an die analogen Verfahren an und nützt die Möglichkeiten einer digitalen Verarbeitung noch nicht aus. Dies gilt auch für das französische CELTIC-System /2.3-5/.

Bei der digitalen Sprachinterpolation (DSI) gibt es im wesentlichen 2 Wege, um auf eine Systemüberlastung zu reagieren und damit die Forderung nach einer konstanten Bitrate auf dem Übertragungskanal zu erfüllen:

- a) Weglassen von einzelnen Abtastwerten
- b) Reduktion der Wortlänge bei den Abtastwerten.

Beim SPEC-System (Speech Predictive Encoding Communications) /2.3-6/ kommt die erste Methode zur Anwendung. Wie bei TASI wird zunächst mit einem Sprachdetektor festgestellt, welche Quellen aktiv sind. Da dies für jeden Abtastwert im Abstand von 125µs neu entschieden wird, kann das SPEC-System sehr viel schneller auf den Sprachbeginn eines Teilnehmers reagieren. Von den aktiven Teilnehmern werden mit zyklischer Priorität 24 ausgewählt und deren Abtastwerte übertragen. Unterdrückte Abtastwerte werden auf der Empfangsseite durch Wiederholung des zuletzt gesendeten Wertes ersetzt. Ein Nachteil ist, daß bei der Entscheidung, welche Abtastwerte nicht übertragen werden sollen, außer den Pausen keine signalabhängigen Kriterien ver-

wendet werden. Dabei ist es möglich, daß Abtastwerte übertragen werden, deren Unterdrückung nur einen geringen Fehler verursachen würde, andere dagegen unterdrückt werden und dabei einen großen Fehler verursachen. Eine verbesserte Version des SPEC berücksichtigt durch eine andere Prioritätsstrategie diese Effekte.

Ebenfalls durch Unterdrückung von Abtastwerten wird beim MPZ-Verfahren (Multiplexverfahren mit momentaner Prioritätszuweisung) /2.3-7/ der Überlastungsfall bewältigt. Dieses Verfahren ist eine Weiterentwicklung eines für Bildsignale entwickelten Datenreduktionsverfahren /2.3-8, 2.3-9, 2.3-10, 2.3-11, 2.3-12/. Das MPZ-System für die Sprachübertragung zeichnet sich vor allem durch die Prioritätsstrategie aus. Es wird ausgenützt, daß man nicht nur Pausenabtastwerte unterdrücken kann, sondern auch aufeinanderfolgende gleiche Abtastwerte, da diese auf der Empfangsseite durch Wiederholen des zuletzt gesendeten Wortes fehlerfrei rekonstruiert werden können. Höchste Übertragungspriorität erhält die Quelle, bei der die größte Differenz zum vorhergehenden Abtastwert auftritt. Sortiert man die Prioritäten der Quellen in absteigender Reihenfolge und überträgt, entsprechend dem Datenreduktionsfaktor, die Abtastwerte der Quellen mit den M höchsten Prioritäten, dann kann man sicher sein, daß der durch die Unterdrückung der restlichen Abtastwerte entstehende Fehler minimal wird. Für das MPZ-System ist kein Sprachdetektor vorgesehen. Der Vollständigkeit wegen soll noch das lagemodulierte Adresscodesystem /2.3-13/ erwähnt werden, das ebenfalls Pausenabtastwerte unterdrückt. Allerdings weist dieses Verfahren bereits ohne Überlastung eine, gegenüber normalen PCM-Systemen, geringere Qualität auf. Die Methode, im Überlastungsfall Abtastwerte zu unterdrücken, führt meist auf relativ einfache Übertragungssysteme. Sie hat jedoch den Nachteil, daß die entstehenden Störungen auf einzelne Quellen konzentriert werden. Vorteilhafter ist ein Verfahren, daß die Störungen möglichst gleichmäßig auf alle Quellen verteilt. Dies ist durch Reduktion der Wortlängen im Überlastungsfall möglich. Bei den italienischen Sprachinterpolationsverfahren VALT (Va-

riable Adaptive Length of Time-slots) /2.3-14/ und PCM-RR (PCM with Rate and Redundancy Reduction) /2.3-15/ wird von diesem Prinzip Gebrauch gemacht. Bei beiden Verfahren werden mit einem Sprachdetektor alle nicht aktiven Quellen von der Übertragung ausgeschlossen. Die aktiven Quellen werden mit 8 bit entsprechend der 13-Segment-Kennlinie codiert. Im Überlastungsfall wird diese Wortlänge reduziert. Durch eine Anpassung an die Signaldynamik kann beim PCM-RR-System die Wortlänge bis auf 3 bit reduziert werden.

In ähnlicher Weise arbeitet das VLC-System (Variable length Coding) /2.3-16/. Auch hier werden im Überlastungsfall niederwertige Bitstellen weggelassen. Ein Sprachdetektor ist bei diesem System nicht vorgesehen; er wird durch eine Einrichtung ersetzt, die in Abschnitten kleiner Amplituden das Signal gleich Null setzt.

Ein anderer Weg wird beim SIDEL-System (Speech Interpolation with Delta Modulation) /2.3-17/ beschritten. Die Rahmenanpassung im Überlastungsfall wird hier durch Variation der Abtastfrequenz bei der Deltamodulation erreicht.

Soweit bekannt, wurde erstmalig beim PWA-Verfahren (Sprachinterpolationsverfahren mit Prädiktionsgesteuerter Wort-Aufteilung) /2.3-18/ das Prinzip der Quellencodierung durch DPCM in ein Sprachinterpolationsverfahren einbezogen.

Das PWA-Verfahren stellt eine Weiterentwicklung des MPZ-Verfahrens dar. Es hat neben der Wortlängenreduktion im Überlastungsfall statt einer Unterdrückung von Abtastwerten vor allem das Ziel, die Adressierungsinformation der aktiven Quellen einzusparen. Das PWA-Verfahren - Gegenstand dieser Arbeit - wird im folgenden ausführlich dargestellt.

Ebenfalls als eine Weiterentwicklung der MPZ- und der PWA-Methode kann das BPZ-Verfahren (Zeitmultiplexverfahren mit Blockweiser Prioritätszuteilung) /2.3-19, 2.3-20/ betrachtet werden. Hier wird durch eine effiziente Quellencodierung (Block- oder Transformationscodierung) die höchste Datenreduktion unter Beibehaltung der Telefoniequalität erreicht.

Eine Kombination von ADPCM mit TASI wird in /2.3-21/ vorge-

schlagen. Als erzielbarer Datenreduktionsfaktor gegenüber PCM ist der Faktor 2 angegeben. In /2.3-22, 2.3-23/ ist ein ausführlicher Vergleich von Sprachinterpolationssystemen, auch hinsichtlich der Übertragungsqualitäten angegeben.

Aus dem bisherigen Rahmen fällt das "Sprachinterpolationsverfahren für nichtöffentliche Netze" /2.3-24/. Hier werden die bei kleinen Quellenzahlen (<16) auftretenden Schwankungen des Übertragungsbedarfs neben der Scharmittelung, zusätzlich durch Zeitmittelung, geglättet. Die dabei auftretenden Verzögerungen von einigen 100 ms sind für private Netze tragbar.

2.4 Neue Beiträge

Schwerpunkt der vorgelegten Arbeit ist die Weiterentwicklung des MPZ-Verfahrens. Durch Anwendung der Quellencodierung (DPCM) wird ein Sprachinterpolationsverfahren entwickelt, das durch unterschiedliche Wortlängen eine bessere Anpassung an die vorhandene Übertragungskapazität ermöglicht und im Überlastungsfall die Störung gleichmäßig auf alle Quellen verteilt. Der Störabstand (SNR) dient dabei als objektives Qualitätskriterium.

Es wird ein Codiervorgang entworfen, das ähnlich der 13-Segment-PCM-Codierung einen vorgegebenen Störabstandsverlauf über dem Signalpegel (Toleranzschema nach CCITT) erfüllt, dabei aber mit einer kleineren Wortlänge auskommt.

Abhängig von den Eigenschaften des Sprachsignals (Dynamik und Korrelation) entsteht ein unregelmäßiger Datenfluß. Das Signal muß mit unterschiedlichen Wortlängen codiert werden. Durch Scharmittelung über eine größere Zahl von Quellen wird der Datenfluß geglättet.

Eine konstante Datenrate auf dem Übertragungskanal, die abhängig vom mittleren Datenfluß der Signalquellen festgesetzt wird, bedingt, daß die momentan benötigte Wortlänge einer Quelle reduziert oder vergrößert werden kann.

Zwei Faktoren sind dafür verantwortlich, daß der Störabstand

zwar im Langzeitmittel über dem CCITT- Toleranzschema liegt, aber bei kurzzeitiger Mittelung streut.

a) das Codierverfahren

(die momentan erforderliche Wortlänge wird aus der Signalvergangenheit geschätzt)

b) Schwankungen durch Scharmittelung über eine endliche Quellenzahl

Beide Einflüsse auf die Störabstandsstreuung sind bei hohen Störabständen (Fehlerrückkopplung hat wenig Einfluß) weitgehend voneinander unabhängig. Deshalb werden zunächst zwei Codierverfahren ohne die Einflüsse der Sprachinterpolation untersucht. Dies entspricht den Verhältnissen in einem Sprachinterpolationssystem mit unendlich vielen Quellen.

Der Einfluß der Sprachinterpolation mit endlicher Quellenzahl wird exemplarisch an dem zweiten Codierverfahren aufgezeigt.

In Kapitel 4.1 wird das PWA1-Verfahren vorgestellt, für das kein Sprachdetektor vorgesehen ist.

Durch den Einsatz eines Sprachdetektors ist es jedoch möglich einerseits die Übertragungsrate zu senken, andererseits noch effizienter arbeitende Quellencodierverfahren anzuwenden. Dies führt zur Entwicklung des in Abschnitt 4.2 beschriebenen PWA2-Verfahrens.

Die Entwicklung beider Sprachinterpolationsmethoden wird durch Computersimulation an einem Kleinrechner (Hewlett Packard 2100) vorgenommen. Ebenfalls Bestandteil dieser Arbeit ist die Entwicklung der dazu notwendigen Ein-/ Ausgabeeinheit am Rechner und die Erstellung der notwendigen System-SOFTWARE (Kapitel 5).

Um einen Überblick über die Realisierungsprobleme eines komplexen Sprachinterpolationsverfahrens zu erhalten, wird in Kapitel 6 ein Signalprozessor vorgestellt, auf dem Teilprobleme des PWA-Verfahrens in Echtzeit realisiert werden können. Daneben wird auf notwendige Rechenzeiten für einige im PWA-Verfahren verwendeten Algorithmen eingegangen.

3. Einführung in das PWA Verfahren

In diesem Abschnitt wird zunächst die Idee skizziert, die dem Sprachinterpolationsverfahren mit prädiktionsgesteuerter Wortaufteilung zugrunde liegt. Für die Bestimmung der momentan notwendigen Wortlänge ist eine genauere Betrachtung des unregelmäßigen Datenflusses bei Sprachsignalen notwendig (Abschnitt 3.2).

Schließlich werden in 3.3 objektive Qualitätskriterien festgelegt, die bei der Systemoptimierung als Gütemaßstab verwendet werden.

3.1 Grundlegende Idee

Aus Erfahrungen mit dem Multiplexverfahren mit momentaner Prioritätszuteilung (MPZ) /2.3-6/ ergaben sich zwei Anhaltspunkte für eine Verbesserung, die zur Entwicklung des PWA-Verfahrens führten.

Beim MPZ-System wird zu jedem Abtastzeitpunkt nur ein Teil der Quellen entsprechend einer Prioritätsstrategie bei der Signalübertragung berücksichtigt. Durch Adressierungsinformation (1 bit/Quelle) muß dem Empfänger mitgeteilt werden, welche Quellen momentan an der Übertragung beteiligt sind. Während die Signale der für die Übertragung ausgewählten Quellen praktisch fehlerfrei übertragen werden, muß bei den unterdrückten Quellen der zuletzt gesendete Wert auf der Empfangsseite wiederholt werden. Durch diese "Alles oder Nichts-Entscheidung" treten zeitlich konzentrierte Störgeräusche auf.

Ziel des PWA-Verfahrens ist es einerseits, die im Überlastungsfall auftretenden Störungen gleichmäßig auf alle Quellen zu verteilen und andererseits die Adressierungsinformation einzusparen. Dies ist möglich, wenn zu jedem Übertragungszeitpunkt alle Quellen in einer festen Übertragungsreihenfolge an

Zentrales Problem bei der Entwicklung der PWA-Methode ist die Bestimmung der für jeden Abtastwert bereitzustellenden Wortlänge. Wie in den nächsten beiden Abschnitten gezeigt wird, ist die Wortlänge von der möglichen Redundanz- und Irrelevanzreduktion abhängig.

Es wird zunächst davon ausgegangen, daß jede Quelle die von ihr angeforderte Wortlänge bekommt.

Erst in einem zweiten Schritt wird die Konkurrenzsituation, bedingt durch Sprachinterpolation bei konstanter Übertragungskapazität, berücksichtigt.

3.2 Der unregelmäßige Datenfluß bei Sprachsignalen

Für die Charakterisierung des Sprachsignals ist es naheliegend, in einem Vergleich mit einem Signal, dessen Eigenschaften bekannt sind, die Unterschiede herauszustellen. Vergleicht man z. B. den Ausschnitt aus einem Sprachsignal mit einem "weißen" Geräusch, so fällt als erstes die größere zeitliche Variation des Sprachpegels auf (Bild 9).

Quantitativ kann diese Eigenschaft dadurch erfaßt werden, daß bei beiden Signalen Segmentleistungen, d.h. über jeweils L Abtastwerte gemittelte Leistungen berechnet werden. Die Streuung S_L der Segmentleistungen um den Wert der mittleren Leistung S_0 hängt von der Segmentlänge ab. Im Anhang wird gezeigt, wiefür das weiße Rauschen als Repräsentant einer statistisch unabhängigen Musterfolge mit Gaußscher AVD, die Streuung abhängig von der Segmentlänge L berechnet werden kann.

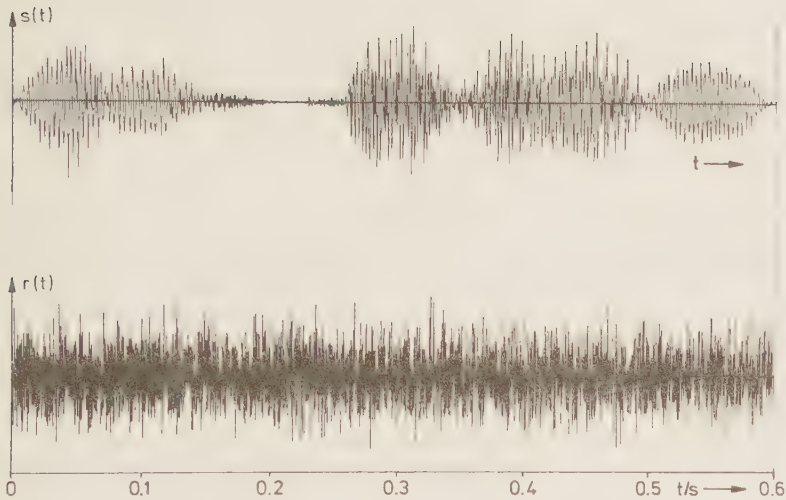


Bild 9: Sprachsignal und weißes Geräusch im Vergleich

Trägt man diesen Zusammenhang in einer graphischen Darstellung (Bild 10) auf, dann kann man mit den durch Messung am Sprachsignal gewonnenen Segmentleistungsstreuungen vergleichen. Aus Vergleichsgründen wurden in den Fällen a), b) und c) nicht die Streuungen direkt, sondern die auf den Mittelwert S_0 der Signalleistung bezogenen Variationskoeffizienten v aufgetragen:

$$v = S_L / S_0$$

Als Ergebnis kann festgestellt werden, daß die Segmentleistung bei Sprachsignalen auch bei großen Segmentlängen noch deutlich stärker variiert, als dies beim Vergleichssignal zutrifft. Diese statistische Abhängigkeit kann benutzt werden, um aus der Signalvergangenheit eine Einteilung des Sprachsignals in Abschnitte unterschiedlicher Signalleistungen vorzunehmen. Bei kleinerer Leistung kann bei gleichem Fehler auch mit kleinerer Wortlänge codiert werden.

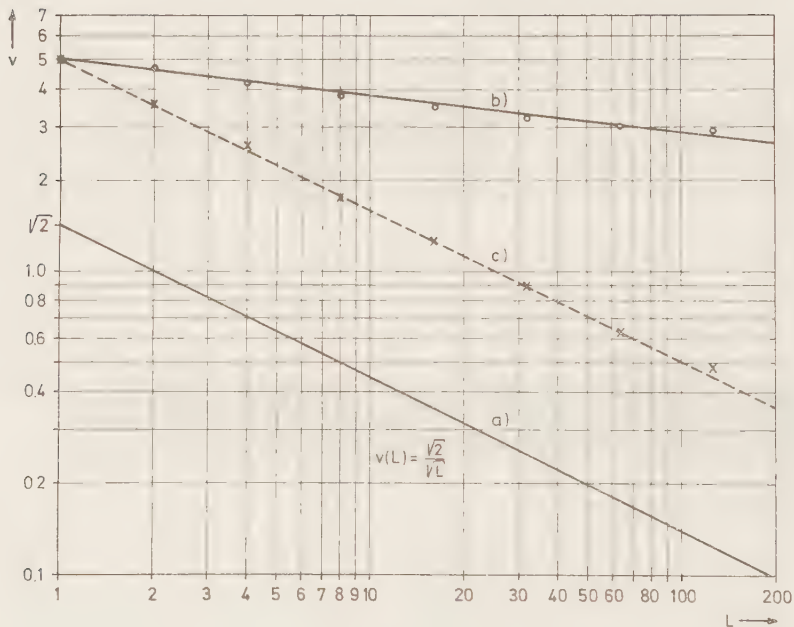


Bild 10: Variationskoeffizient v der Segmentleistungen abhängig von der Segmentlänge L für

- a) weißes Rauschen mit gaußscher AVD
- b) Sprachsignal
- c) Sprachsignal, jedoch Segmentbildung nicht durch zeitliche, sondern durch Scharmittelung

Ein anderes Ergebnis mit dieser Betrachtungsweise bestätigt die Vorteile einer Scharmittelung gegenüber einer Zeitmittelung bei Sprachsignalen.

Bei der unterbrochenen Linie c) in Bild 10 wurde gegenüber der Linie a) die Segmentleistung nicht durch zeitliche Mittelung über L Abtastwerte einer Quelle, sondern durch Scharmittelung über L Quellen berechnet. Eine Reduktion der Streuung gegenüber $L=1$ um den Faktor 2 ist entweder durch Scharmittelung über 4 Quellen, oder durch Zeitmittelung über 300 Abtastwerte ($= 125 \text{ ms}$) zu erreichen.

Gegenüber 5 Quellen bei Scharmittelung müssen durch Zeitmittelung bereits 1000 Abtastwerte gemittelt werden, um die gleiche Reduktion der Streuung zu erhalten.

Neben einer "makroskopischen" Betrachtung des Sprachsignals, wo der zeitlich schwankende Signalpegel in Erscheinung tritt, kann man auch bei feinerer zeitlicher Auflösung eine bereichsweise Unterscheidung von Signaleigenschaften vornehmen. Ein dekorreliertes Signal hat ein weißes Spektrum. Wenn die Spektrallinien dagegen in einem Spektrum konzentriert auftreten, kann auf eine mehr oder weniger große Korrelation im Signal geschlossen werden (Bild 11).

Dies trifft vorallem auf Vokale oder auf Nasallaute wie /m/ und /n/ zu. Hier ist der Einsatz von DPCM-Verfahren vorteilhaft, da dann durch den Prädiktionsgewinn mit einer um 1 bis 2 bit kürzeren Wortlänge codiert werden kann. Im gleichen Bild sind aber auch Signalabschnitte erkennbar, die kaum noch Redundanz enthalten. Es handelt sich dabei um Zischlaute. Dies wird besonders bei Betrachtung des Kurzzeitspektrums deutlich. Beim Einsatz von DPCM-Verfahren scheidet häufig aus Aufwandsgründen eine Adaption der Prädiktionskoeffizienten aus. Aus der Langzeit-Autokorrelationsfunktion von Sprache wird dann ein "optimaler" Koeffizientensatz berechnet /2.2-6/.

Da im Sprachsignal die Abschnitte hoher Korrelation meist in Verbindung mit einer sehr hohen Signalleistung auftreten, dominieren diese Abschnitte gegenüber den Bereichen niederer Korrelation mit sehr viel kleinerer Leistung bei der Berechnung der Prädiktionskoeffizienten.

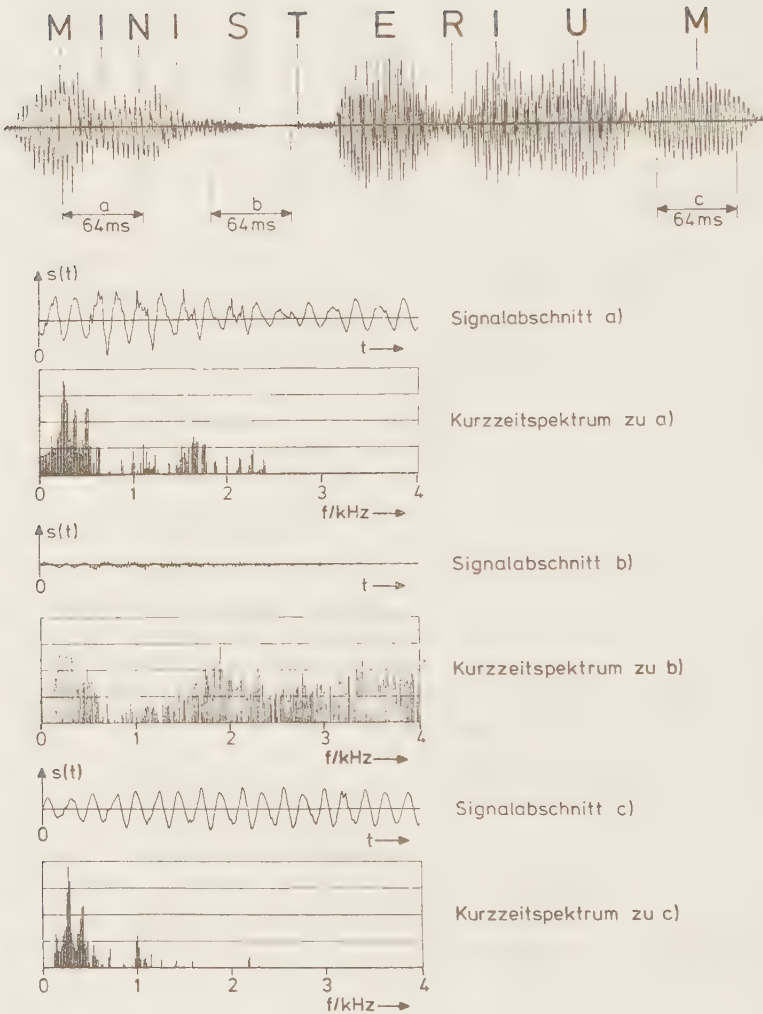


Bild 11: Unterschiedliche Korrelation im Sprachsignal
 (Kurzzeitspektrum von je 512 Abtastwerten, bewertet
 mit Rechteckfenster)

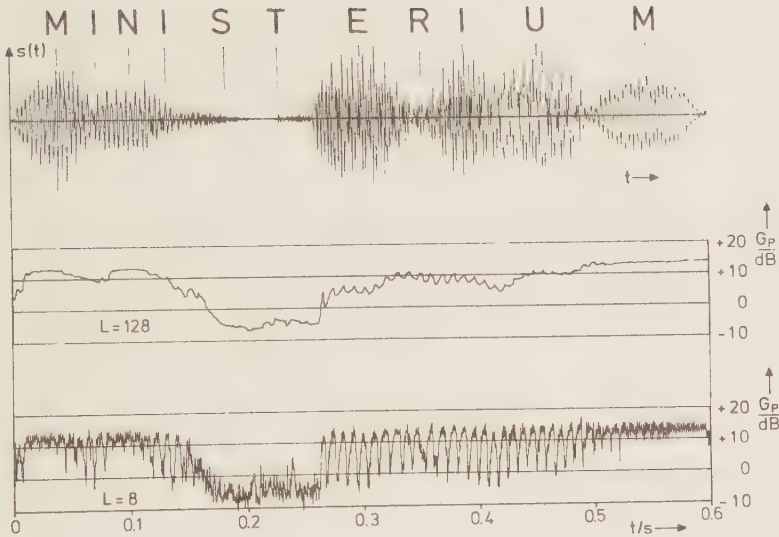


Bild 12: Zeitlicher Verlauf des Prädiktionsgewinns G_p für unterschiedliche Segmentlängen L

Dies führt dann oft dazu, daß die Prädiktion wenig korrelierter Abschnitte nicht nur schlecht gelingt, sondern auch falsch werden kann. In diesen Fällen wird die Leistung der DPCM-Werte größer als die Leistung der PCM-Werte σ_s^2 und der Prädiktionsgewinn G_p negativ.

$$G_p = 20 \lg \frac{\sigma_s}{\sigma_p}$$

Es erscheint deshalb sinnvoll, während dieser Signalabschnitte von einer DPCM- auf eine PCM-Übertragung umzuschalten. In Bild 12 ist der zeitliche Verlauf des Prädiktionsgewinns aufgetragen.

Der mittlere Prädiktionsgewinn beträgt 7.5 dB.

Bild 13 zeigt die Summenhäufigkeit von Prädiktionsgewinnen bei Verwendung eines Prädiktors 2. Grades mit festen Prädiktorkoeffizienten. Der gewählte Koeffizientensatz ($a_1 = 1.25$, $a_2 = -0.5$) stellt für eine große Zahl von Sprachproben einen guten Kompromiß dar. In ca. 20 % der Fälle ist eine PCM-Übertragung günstiger als eine DPCM-Übertragung.

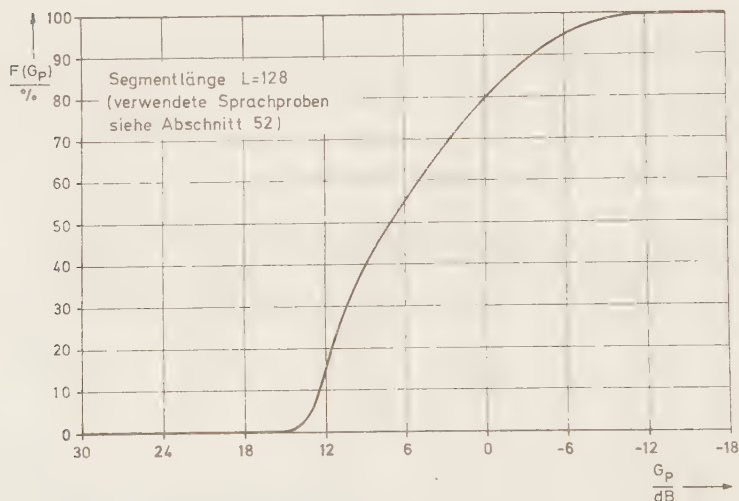


Bild 13: Summenhäufigkeit F von Prädiktionsgewinnen

Bisher nicht erwähnt wurden die Pausenabschnitte im Sprachsignal, für die im Idealfall keine Übertragungskapazität zur Verfügung gestellt werden sollte.

Die qualitativen Betrachtungen über den unregelmäßigen Datenfluß werden in den Abschnitten 4.1.4 und 4.2.5 durch quantitative Aussagen bezüglich der notwendigen Wortlängen bei der Codierung ergänzt.

3.3 Festlegung von Qualitätskriterien

In der Aufgabenstellung, auf der diese Arbeit aufbaut, war als wichtiger Bestandteil die Verwendung von objektiven Qualitätskriterien festgelegt. Dies bedeutet im vorliegenden Fall die Verwendung des Störabstandes (SNR).

In Anlehnung an die 13-Segment-PCM-Codierung wurde das Toleranzschema nach CCITT /3.3-1/, das über einen Dynamikbereich von mehr als 50 dB einen Mindeststörabstand festlegt, als Grundlage für die Systemoptimierung ausgewählt (Bild 14). Die Erfüllung des Toleranzschemas stellt an das Codierverfahren sehr hohe Anforderungen. Ob diese gerechtfertigt sind, soll hier nicht näher diskutiert werden. Es hat sich gezeigt, daß auch bei ca. 4-6 dB niedrigeren Störabständen noch keine hörbaren Störungen auftreten. Dies gilt um so mehr, je stärker subjektive Qualitätskriterien berücksichtigt werden. Nachdem jedoch die Einführung der 13-Segment PCM feststeht, wird dieser "Standard" bis zur Bereitstellung besserer objektiver Meßmethoden bei der Entwicklung neuer Quellcodierverfahren berücksichtigt werden müssen. Zur Berechnung des Störabstandes über dem Signalpegel wird für Segmente von z.B. 128 Abtastwerten der Signalpegel und der Störabstand ermittelt. Trägt man diese Segmentergebnisse in das Toleranzschema ein, so erhält man eine Punktwolke, die zunächst nicht sehr aussagekräftig ist. Nach /2.3-20/ kann man jedoch eine Mittelwertkurve durch Berechnung der mittleren Fehlerleistungen in einzelnen Signalpegelintervallen (z B. Intervalle von 1 dB Breite) angeben. Dies ist für ein Simulationsergebnis in Bild 14 eingetragen. Als Beispiel für die Codierung wurde die bekannte 13-Segment - PCM verwendet. Sie wird später noch zum Vergleich mit Simulationsergebnissen beim PWA-Verfahren verwendet.

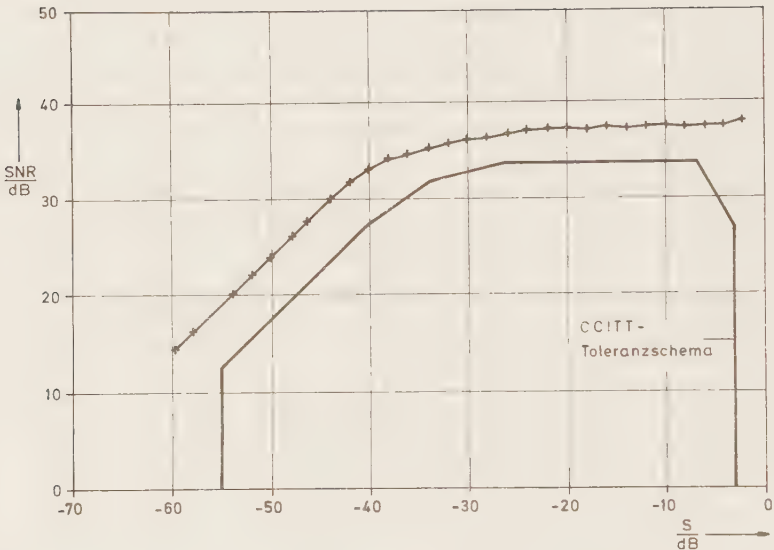


Bild 14: Toleranzschema nach CCITT mit Simulations-
ergebnis einer 13-Segment-PCM-Codierung (8 bit).

Eine übersichtliche Darstellung der Störabstandsstreuung erhält man, wenn nur noch der Abstand ΔSNR der Segment-SNR-Werte zum Toleranzschema betrachtet wird. Man erhält einen mittleren Abstand zur Toleranzgrenze und die dazugehörige Streuung in dB. In einer Summenverteilung aufgetragen, vermittelt dieser SNR-Abstand eine gute Vergleichsmöglichkeit verschiedener Codierverfahren.

Zur Demonstration dieser Darstellung wird wieder die 13-Segment-Codierung verwendet. Bei einer Wortlänge von 8 bit beträgt der mittlere Abstand zur Toleranzgrenze +4.05 dB und die $1\text{-}\sigma$ -Streuung 2.45 dB.

Die Summenverteilung der SNR-Abstände zum Toleranzschema ist in Bild 15 für die 13-Segment-Codierung für 8 bit und 7 bit angegeben.

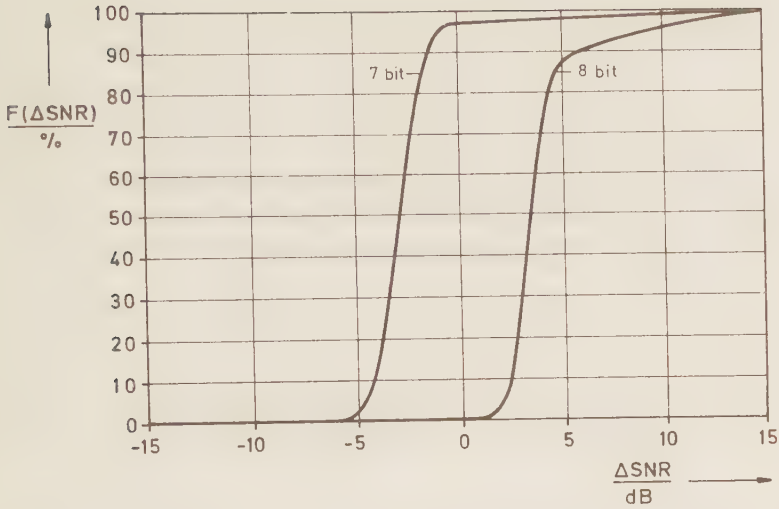


Bild 15: Summenhäufigkeit der SNR-Abstände zum Toleranzschema ΔSNR für 13-Segment-Codierung mit unterschiedlichen Wortlängen w

4. Die Anwendung des PWA-Prinzips

In diesem Abschnitt werden zwei mögliche Ausführungsformen des PWA-Verfahrens aufgezeigt. Kapitel 4.1 beschreibt ein einfaches System, bei dem kein Sprachdetektor vorgesehen ist. In Abschnitt 4.2 wird gezeigt, welche Verbesserungen durch den Einsatz eines Sprachdetektors möglich sind.

4.1 PWA1-Sprachinterpolation mit prädiktions- gesteuerter Wortaufteilung ohne Sprachdetektor

Bei Vorarbeiten zu diesem System konnte in /2.3-16/ gezeigt werden, daß die PWA-Methode auch bei häufiger Überlastung stabil arbeitet.

In den nachfolgenden Untersuchungen galt es, für den zeitlichen Verlauf des Sprachsignals die momentan erforderliche Wortlänge zu bestimmen.

Dies gelingt einerseits durch eine Klassifizierung des augenblicklichen Signalzustandes und andererseits durch Angabe eines zulässigen Fehlers.

Der zulässige Fehler kann mit dem Toleranzschema nach CCITT in Abhängigkeit vom Signalpegel bestimmt werden. Es ist deshalb naheliegend, das Sprachsignal nach dem momentanen Signalpegel zu klassifizieren.

Für jeden dieser Signalebereiche kann die bedingte Häufigkeitsverteilung der Signalmomentanwerte (kurz PCM-Werte) gemessen werden und eine näherungsweise optimale Quantisierungskennlinie berechnet werden. Dabei wird durch die vom Toleranzschema zugelassene Fehlerleistung die notwendige Wortlänge bei der Codierung festgelegt.

Allerdings ist noch nicht die Redundanz durch Korrelation im Sprachsignal berücksichtigt. Durch den Einsatz der dekorrelierenden Differenzpulsmodulation stehen neben den PCM-Werten

auch noch die DPCM-Fehlersignale (kurz: DPCM-Werte) zur Verfügung, mit deren Hilfe eine weitere Klassifizierung des Sprachsignals möglich ist. Für diese unterschiedlichen Signalklassen können über die Häufigkeitsverteilungen der PCM- und der DPCM-Werte, die PCM- und DPCM-Leistungen, die Prädiktionsgewinne und die Quantisierungskennlinien berechnet werden.

Diese Einteilung ermöglicht, abhängig vom Prädiktionsgewinn einer Gruppe, eine einfache Entscheidung, ob DPCM- oder PCM-Werte bei gleichem Fehler mit der kleineren Wortlänge übertragen werden können.

4.1.1. Klassifizierung des Sprachsignals

Darf die Signalklassifizierung wie im vorliegenden Fall nur mit Größen aus der Signalvergangenheit erfolgen, dann muß an Stelle der momentanen Amplitude eine andere Kenngröße gefunden werden.

Hier bietet sich die Einhüllende des Sprachsignals an. Da sich diese Kurve verhältnismäßig langsam ändert, kann mit ihr gut von einem Abtastzeitpunkt zum nächsten auf die wahrscheinliche Signalamplitude geschlossen werden. Dabei ist es ausreichend, die Einhüllende nicht exakt, sondern nur näherungsweise zu berechnen ($E(k)$).

Die einfachste Näherung ($E_1(k)$) ergibt sich durch lineare Mittelung über die Beträge von vergangenen Signalmomentanwerten. Die Länge des verwendeten Signalsegments soll mit L bezeichnet werden:

$$E_1(k) = \frac{1}{L} \sum_{i=1}^L |u(k-i)|$$

Eine andere Näherung erhält man durch quadratische Mittelung (Kurzzeiteffektivwert):

$$E_2(k) = \sqrt{\frac{1}{L} \sum_{i=1}^L u^2(k-i)}$$

Der Einfluß der Segmentlänge L auf beide Näherungen sei an einem einfachen Beispiel erläutert. Als Testsignal wird eine Sinusschwingung mit einer Frequenz von 300 Hz ein- und ausgeschaltet. Die Einhüllende hat in diesem Beispiel einen rechteckförmigen Verlauf. In Bild 16 sind neben der geschalteten Sinusschwingung Näherungen für die Einhüllende aufgetragen.

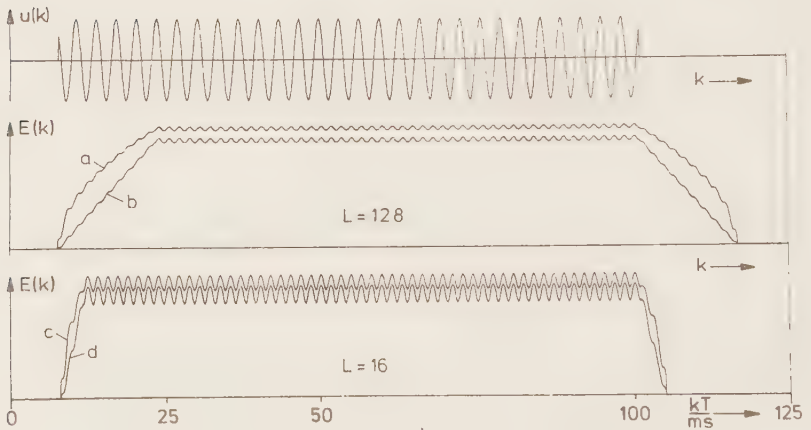


Bild 16: Näherung für die Einhüllende $E(k)$ einer geschalteten Sinusschwingung
 a) und c) : quadratische Mittelung
 b) und d) : lineare Mittelung

Die Kurven a) und c) wurden durch quadratische, die Kurven b) und d) durch lineare Mittelung gewonnen. Die Segmentlänge, mit der die Mitteldauer festgelegt wird, beträgt für a) und b) 128 Abtastwerte, für c) und d) 16 Abtastwerte.

Abgesehen vom linearen Verlauf des Anstiegs und des Abfalls bei der linearen Mittelung, im Gegensatz zum Verlauf nach einer Wurzelfunktion bei der quadratischen Mittelung, fällt vor allem

die Abhängigkeit von der Segmentlänge auf. Große Segmentlänge bedeutet lange Anstiegs- und Abfallszeiten und geringe Schwankungen im "stationären" Bereich. Eine kürzere Anstiegszeit muß durch größere Schwankungen erkaufte werden.

Um mit diesen Näherungen an die Einhüllende eine Einteilung des Sprachsignals in Bereiche unterschiedlicher Signalamplituden und damit die Verwendung verschiedener Quantisierungskennlinien zu ermöglichen, muß ein Kompromiß bezüglich der Segmentlänge getroffen werden.

Die optimale Segmentlänge wurde durch Simulation mit verschiedenen Segmentlängen ermittelt. Als Ergebnis dieser Untersuchung erhält man bei der Mittelung von PCM-Werten ein flaches Optimum für Segmentlängen zwischen 8 und 16 Abtastwerten. Für die Mittelung von DPCM-Werten ist ein flaches Optimum bei Segmentlängen zwischen 4 und 8 Abtastwerten vorhanden. Die quadratische Mittelung liefert geringfügig bessere Ergebnisse als die lineare. In Anbetracht der einfacheren Realisierung wird in einem realen System jedoch die lineare Mittelung den Vorrang erhalten. Als Optimierungskriterium wurde der mittlere Störabstand verwendet.

Für die folgenden Ausführungen ist eine Segmentlänge L_s von 16 PCM-Werten und L_d von 4 DPCM-Werten zugrunde gelegt. Es wird die quadratische Mittelung verwendet.

Die eigentliche Signalklassifizierung erfolgt zunächst nach dem Schätzwert für die Einhüllende der PCM-Werte, der mit $S^*(k)$ bezeichnet werden soll. Je nach Größe von S^* wird das Sprachsignal einer bestimmten Signalpegelklasse zugeordnet. Als nächstes muß die Frage nach der Klassenzahl für die Signalpegel und deren Grenzen diskutiert werden.

Die Häufigkeitsverteilung der Signalpegel in Bild 17 weist zwei Maxima aus: ein Maximum bei kleinen Signalpegeln, das andere bei großen Signalpegeln.

Es liegt nun nahe, diese beiden Signalbereiche abzuspalten, und dem verbleibenden Rest eine weitere Klasse zuzuordnen. Dies gelingt dann ohne weiteres, wenn man für die kleinen und großen Signalpegel vereinfachend symmetrische Verteilungen annimmt. Bild 18 zeigt diese Aufspaltung.

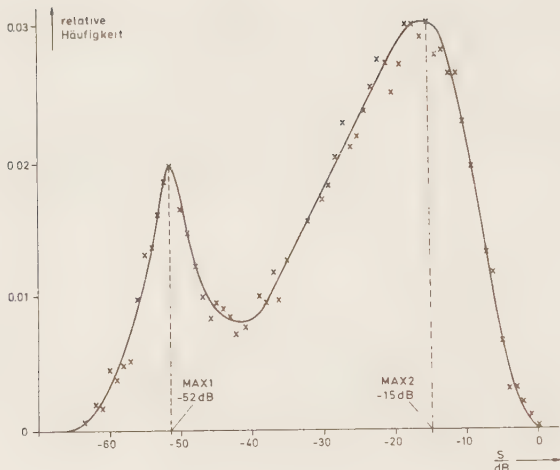


Bild 17: Relative Häufigkeiten des Signalpegels
(quadratische Mittelung über 16 Abtastwerte,
verwendete Sprachproben siehe Abschnitt 5.2)

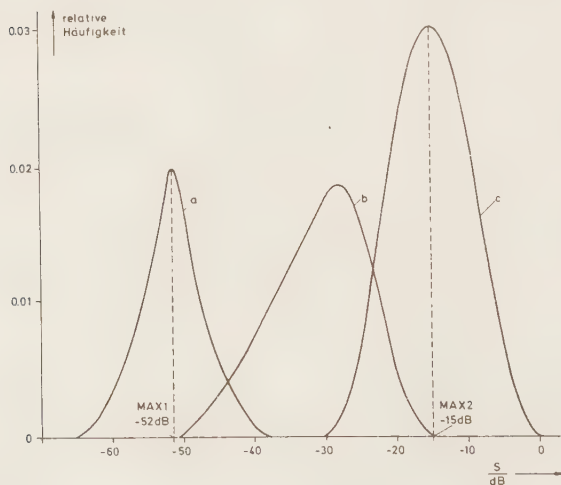


Bild 18: Abspaltung von zwei symmetrischen Verteilungen
a) für kleine, c) für große Pegel. b): Restverteilung

Berechnet man zu der Signalpegelverteilung in Bild 17 die Summenverteilung (Bild 19), so kann durch die Annahme der Symmetrie für die Verteilung der kleinen und großen Signalpegel leicht der prozentuale Anteil für diese beiden Bereiche ermittelt werden.

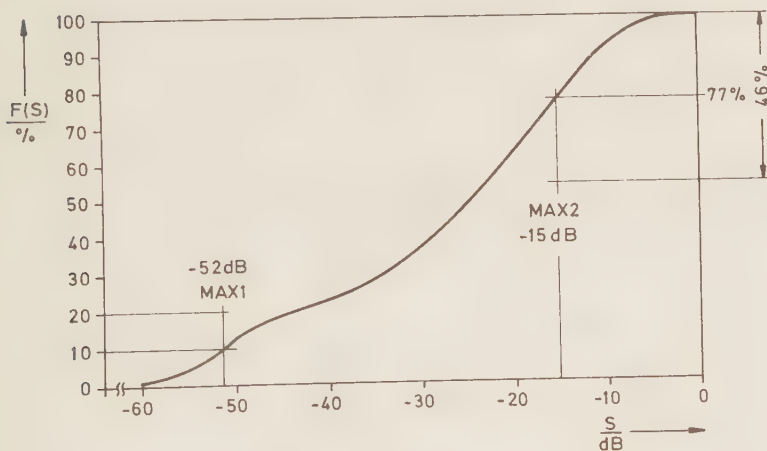


Bild 19: Summenverteilung der Signalpegel

Für Sprache ohne lange Pausen gilt:

Klasse		Anteil
kleine Pegel	KS1	20 %
große Pegel	KS3	46 %
Rest	KS2	34 %

Im nächsten Schritt müssen die Entscheidungsschwellen für den PCM-Schätzwert S^* bestimmt werden.

Für die Klasse der kleinen Pegel KS1 soll gelten:

$$0 \leq S^* \leq \text{SKS1}$$

Sinngemäß gilt für die Klasse der großen Pegel:

$$\text{SKS3} \leq S^* \leq 2048$$

In Bild 20 sind für die Klassen KS1 und KS3 Ausschnitte aus den Summenhäufigkeiten mit den Parametern SKS1 und SKS3 dargestellt.

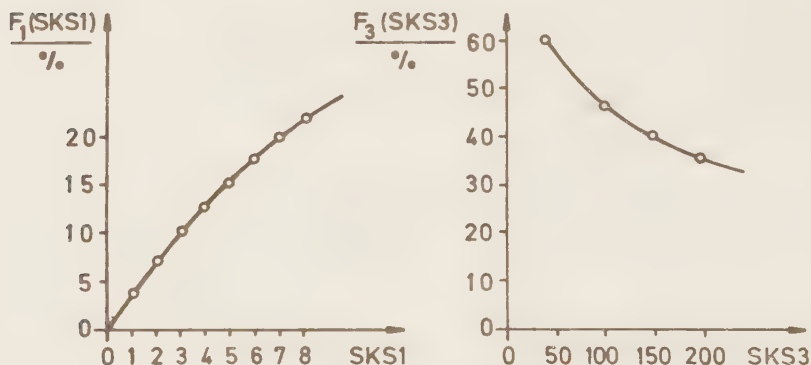


Bild 20: Ausschnitte aus Summenverteilungen der Klassen KS1 und KS3 abhängig von der Entscheidungsschwelle SKS1 und SKS3

Nachdem die Klasse KS1 einen Anteil von 20 % haben soll, kann für die Entscheidungsschwelle SKS1 der Wert 7 abgelesen werden. Die Klasse KS3 hat bei $\text{SKS3} = 100$ den gewünschten Anteil von 46 %.

Damit ergibt sich automatisch für die Restklasse KS2:

$$\text{SKS1} < S^* < \text{SKS3}$$

Für die weitere Klassifizierung des Sprachsignals durch Aufteilung jeder Signalpegelklasse in Untergruppen, wird der Schätzwert für die Einhüllende der DPCM-Werte herangezogen,

der im folgenden mit $D^*(k)$ bezeichnet werden soll. $D^*(k)$ wird durch quadratische Mittelung über 4 zurückliegende DPCM-Fehlersignale berechnet. Durch diese Unterscheidung nach großen oder kleinen DPCM-Werten, kann der unterschiedliche Prädiktionsgewinn ausgenutzt werden. Kriterium für die Zahl und Größe der Entscheidungsschwellen ist die Abstufung der mittleren DPCM-Leistung jeder Untergruppe in 6 db - Schritten. Auch hier hat sich eine Einteilung in 3 Gruppen (kleine, mittlere und große DPCM-Werte) als guter Kompromiß erwiesen.

Die Aufteilung jeder Signalpegelklasse in 3 Untergruppen wird durch die Entscheidungsschwellen der mittleren Gruppe eindeutig festgelegt. Es gilt:

in KS1:	$SG11 < D^* < SG13$	mit $SG11=2$; $SG13=4$
in KS2:	$SG21 < D^* < SG23$	mit $SG21=10$; $SG23=40$
in KS3:	$SG31 < D^* < SG33$	mit $SG31=40$; $SG33=100$

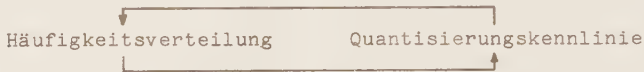
Die so entstandenen 9 Gruppen sind in einer schematischen Darstellung in Bild 21 zusammengefaßt.

			PCM - Werte
kleine	mittl.	große	
KS1	KS2	KS3	
G 11	G 21	G 31	kleine DPCM - Werte
G 12	G 22	G 32	mittlere DPCM - Werte
G 13	G 23	G 33	große DPCM - Werte

Bild 21: Klassifizierung des Sprachsignals

4.1.2 Bestimmung der bedingten Häufigkeitsverteilungen für PCM- und DPCM-Werte in den 9 Gruppen

Abhängig vom PCM-Schätzwert S^* und vom DPCM-Schätzwert D^* kann für jede der 9 Gruppen die bedingte Häufigkeitsverteilung der PCM- und der DPCM-Werte gemessen werden. Diese Verteilungen stellen die Grundlage für die Berechnung der optimalen Quantisierungskennlinien für jede Gruppe dar. Die daraus resultierenden Quantisierungsfehler beeinflussen aber wieder durch die Fehlerrückkopplung des DPCM-Systems die einzelnen Häufigkeitsverteilungen. Streng genommen ist die Ermittlung der Häufigkeitsverteilungen nur in einem iterativen Prozeß möglich:



Glücklicherweise konvergiert dieser Prozeß bei kleinen Quantisierungsfehlern ($\text{SNR} > 25 \text{ dB}$) sehr rasch. Bereits nach dem ersten Durchlauf ist kaum noch eine Änderung feststellbar. Bild 22 gibt einen Einblick in die entstehenden Häufigkeitsverteilungen. Als Beispiel sind die 9 Verteilungen der DPCM-Werte dargestellt.

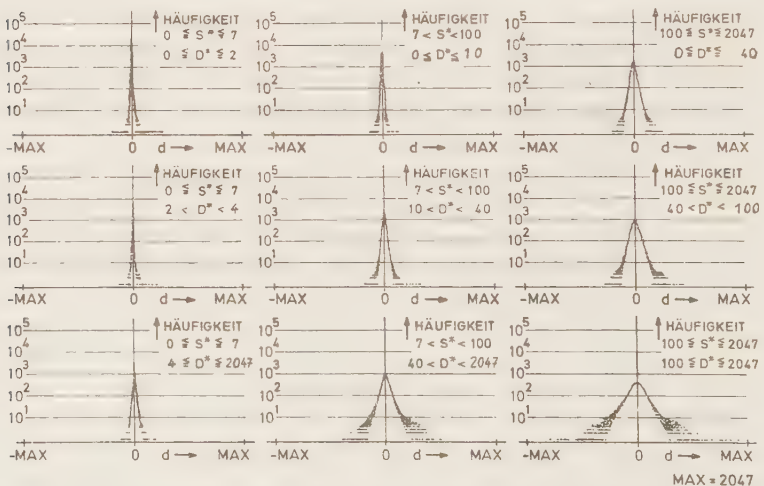


Bild 22: Häufigkeitsverteilungen der DPCM-Werte $d(k)$
in den 9 Gruppen

Die Messung der Häufigkeitsverteilungen wurde unter Verwendung verschiedener Sprachproben durchgeführt, um einigermaßen repräsentative Ergebnisse zu erhalten.

4.1.3 Berechnung von optimalen Quantisierungskennlinien

Für die ermittelten Häufigkeitsverteilungen sollen Quantisierungskennlinien berechnet werden, die bei gegebener Verteilung eine minimale Quantisierungsfehlerleistung gewährleisten. Eine Lösung für dieses Problem wurde von MAX /4.1-1/ angegeben. Allerdings ist, mit Ausnahme einiger Spezialfälle, keine geschlossene Lösung möglich. Die Quantisierungskennlinien können nur iterativ berechnet werden. Weiterhin wird von einem Amplitudenkontinuum ausgegangen, es sind für die Quantisierungsstufenbreiten und deren Repräsentanten beliebige reelle Zahlen zugelassen.

Gerade diese Voraussetzung entfällt aber in der vorliegenden Aufgabenstellung. Durch die lineare Vorquantisierung mit 12 bit sind nur noch 4096 diskrete Amplitudenwerte vorhanden. An Stelle einer Wahrscheinlichkeitsdichteverteilung $p(u)$ steht eine Verteilung der relativen Häufigkeiten $p(u_i)$ zur Verfügung. Die berechneten Quantisierungsstufenbreiten und die Repräsentanten sollen wieder Elemente der ganzen Zahlen sein. Prinzipiell wäre es möglich, durch Interpolation zwischen den diskreten Werten das Problem auf einen kontinuierlichen Zahlenbereich zu übertragen. Aber spätestens nach der Berechnung der Kennlinie müßten die Repräsentanten und die Grenzen der Quantisierungsstufen wieder auf die ganzen Zahlen abgebildet werden. Die dabei auftretende Abweichung von der minimalen Quantisierungsfehlerleistung kann, wie bei eigenen Untersuchungen beobachtet wurde, größer sein, als wenn die Optimierung von vornherein im diskreten Bereich vorgenommen wird.

Der im folgenden beschriebene Vorschlag für die Berechnung von

näherungsweise optimalen Kennlinien im Bereich diskreter Amplituden liefert einen schnell konvergierenden Iterationsalgorithmus für einen Digitalrechner, der als Eingabeparameter nur die Verteilung der relativen Häufigkeiten und die Zahl der Quantisierungsstufen benötigt.

Normalerweise sind die verwendeten Häufigkeitsverteilungen symmetrisch. Dadurch ist es möglich, eine Darstellung nach Betrag und Vorzeichen zu wählen. Für die Quantisierung wird dann nur noch der Zahlenbereich zwischen 0 und 2047 relevant (2048 wird bei der Betragsbildung auf +2047 "abgerundet"). Eine Einteilung dieses Bereichs in N Quantisierungsintervalle wird durch N-1 Intervallgrenzen eindeutig festgelegt (Bild 23).



Bild 23: Quantisierung des Zahlenbereichs

Für jedes Quantisierungsintervall gibt es einen Repräsentanten r_v , der so gewählt werden muß, daß der Quantisierungsfehler $\sqrt{Q_v}$ im Intervall minimal wird:

$$Q_v(r_v) = \sum_{i=I_{v-1}}^{I_v} (u_i - r_v)^2 \cdot p(u_i) \quad \text{mit } u_i = 0, 1, \dots, 2047 \\ p(u_i), r_v \text{ reell}$$

$$Q_{v \min}: \frac{dQ_v(r_v)}{dr_v} = 0 \Rightarrow \sum_{i=I_{v-1}}^{I_v} 2(u_i - r_v) \cdot p(u_i) = 0$$

$$r_v = \frac{\sum_{i=I_{v-1}}^{I_v} u_i \cdot p(u_i)}{\sum_{i=I_{v-1}}^{I_v} p(u_i)}$$

Da r normalerweise nicht ganzzahlig ist, muß ein genäherter ganzzahliger Repräsentant K bestimmt werden. Hier kann man die Kenntnisse über die Häufigkeitsverteilung benutzen, um K festzulegen. Die relativen Häufigkeiten $p(u_i)$ fallen zu größeren Amplituden ab:

$$p(u_i) > p(u_{i+1})$$

Es gilt deshalb:

$$r_{v-1} < K_v \leq r_v \quad \text{mit } r_v \text{ reell; } K_v \text{ ganzzahlig}$$

Optimierungsansatz ist, die gesamte Quantisierungsfehlerleistung Q_0 minimal zu machen:

$$Q_0 = \sum_{v=1}^N Q_v \rightarrow \text{minimal}$$

Untersuchungen haben gezeigt, daß Q_0 gerade dann minimal wird, wenn alle Q_v möglichst gleich groß sind.

Eine Einteilung nach dieser Annahme ermöglicht in einfacher Weise die Nachprüfung dieser Hypothese. Betrachtet man nämlich jeweils zwei nebeneinanderliegende Quantisierungsintervalle und variiert die mittlere Intervallgrenze, so darf die Summe der beiden Quantisierungsfehlerleistungen bei Richtigkeit der Hypothese nicht kleiner werden.

Für die untersuchten Häufigkeitsverteilungen konnte die Hypothese bestätigt werden.

In der Literatur /4.1-2/ ist für den kontinuierlichen Zahlenbereich der Amplituden eine analytische Lösung angegeben, die mit der vereinfachenden Annahme, daß die Repräsentanten genau in der Intervallmitte liegen, zu dem gleichen Ergebnis führt. Es ist deshalb nicht verwunderlich, daß auch im vorliegenden Fall (in Analogie zu /4.1-2/) ein Zusammenhang zwischen unterschiedlichen Intervallquantisierungsfehlerleistungen und zugehörigen Intervallzahlen angegeben werden kann. Q_{v1} soll auf N_1 Intervalle, Q_{v2} auf N_2 Intervalle führen. Es gilt mit guter Näherung:

$$\frac{Q_{v1}}{Q_{v2}} = \left(\frac{N_2}{N_1} \right)^3$$

In Bild 24 ist in einer doppeltlogarithmischen Darstellung dieser Zusammenhang für 3 willkürlich ausgewählte Häufigkeitsverteilungen angegeben.

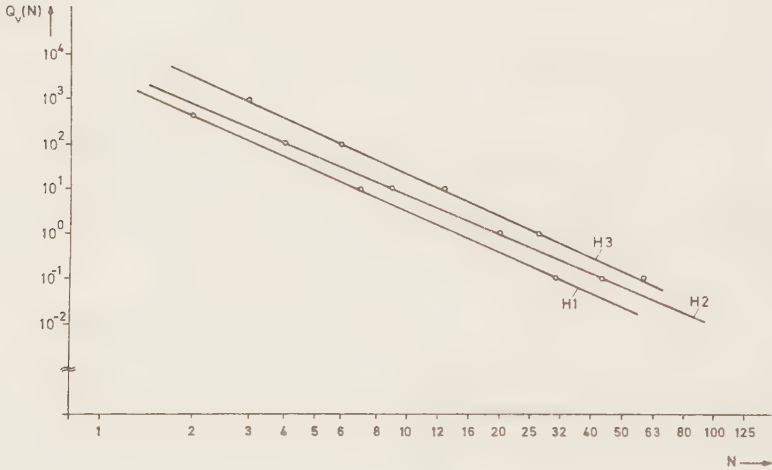


Bild 24: Optimale Intervallquantisierungsfehlerleistung Q_I abhängig von der Intervallzahl N für verschiedene Häufigkeitsverteilungen

Mit der Vorgabe einer bestimmten Intervallfehlerleistung Q_I ist durch die Annahme, daß die Fehlerleistung in allen Intervallen gleich sein soll, eine einfache Bestimmung der Intervallgrenzen I_v möglich.

Man beginnt mit dem ersten Intervall. Die Quantisierungsfehlerleistung ist dann nur noch von I_1 abhängig. Es wird der Wert von I_1 ausgewählt, dessen zugeordnete Intervallfehlerleistung am nächsten an Q_I liegt.

Dies kann dann für das nächste und für alle weiteren Intervalle in gleicher Weise durchgeführt werden.

Ausgehend von einem nahezu beliebigen Wert von Q_{I_x} erhält man eine bestimmte Intervallzahl N_x . Diese wird sicher nicht mit der gewünschten Zahl N übereinstimmen. Die Iteration erfolgt

durch Berechnung eines verbesserten Startwertes:

$$Q'_I = Q_{I_*} \cdot \left(\frac{N_x}{N} \right)^3$$

der häufig bereits zur richtigen Zahl von Quantisierungsstufen führt.

Die Optimierung wird nach Erreichen des richtigen Wertes von N dadurch abgeschlossen, daß durch weitere Iteration die kleinstmögliche Intervallfehlerleistung $Q_{\min}$ bestimmt wird, für die gerade noch N Intervalle möglich sind.

Es hat sich gezeigt, daß auch bei Startwerten, die um mehrere Zehnerpotenzen neben dem richtigen Wert liegen, noch eine rasche Konvergenz erreicht wird.

Die gemessenen Häufigkeitsverteilungen weisen im Bereich großer Amplituden (durch die geringe Auftrittswahrscheinlichkeit) eine verhältnismäßig große Streuung auf. Im Interesse einer guten Konvergenz ist eine Glättung zweckmäßig. Hierfür wurde ein Vorschlag aus /4.1-3/ angewendet.

4.1.4 Codierung der Einzelquelle

Zunächst soll nur der Einfluß der gewählten Quellencodierung auf den Störabstand untersucht werden. Es wird angenommen, daß die angeforderte Wortlänge einer Quelle (ohne Beeinflussung durch Sprachinterpolation) auch zugeteilt wird. Dies entspricht den Verhältnissen in einem Sprachinterpolationssystem mit unendlich vielen Quellen.

Nach der Berechnung von optimalen Quantisierungskennlinien kann für jeden der 9 Signalbereiche die Quantisierungsfehlerleistung und mit Hilfe der relativen Häufigkeitsverteilungen

der PCM-Signale die mittlere Signalleistung angegeben werden. Daraus kann für jeden Bereich ein Störabstand SNR berechnet werden.

Trägt man diese Störabstände bei verschiedenen Bitzahlen über den dazugehörenden Signalpegeln auf, dann kann man mit dem CCITT-Toleranzschema sofort entscheiden, mit welcher Wortlänge jeder Bereich codiert werden muß. Bild 25 zeigt diesen Sachverhalt. Die zur Erfüllung des Toleranzschemas erforderlichen Wortlängen sind besonders gekennzeichnet. Zusätzlich ist der SNR-Verlauf über dem Signalpegel eines Simulationsergebnisses eingetragen, bei dem diese Wortlängen verwendet wurden.

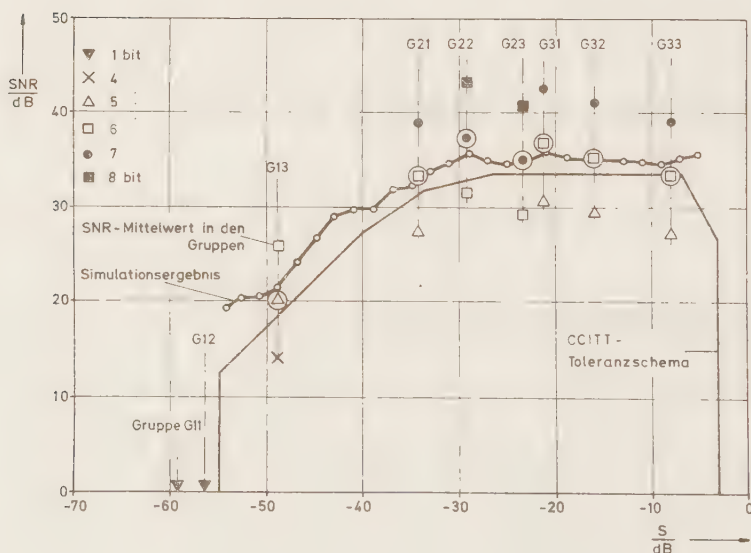


Bild 25: SNR-Werte der 9 Signalbereiche abhängig von der Wortlänge und ein Simulationsergebnis SNR über S

In einer Rechnersimulation wurde festgestellt, daß bei durchgehend gesprochenem Text je nach Sprachprobe zwischen 4.6 und 5.6 bit pro Abtastwert aufgewendet werden muß, um mit dem entwickelten Codierverfahren das Toleranzschema zu erfüllen. Der mittlere Störabstand beträgt dann 34 bis 36 dB.

Entsprechend einem Pausenanteil von ca. 13 % werden die Bereiche G11 und G12 mit der Wortlänge 1 bit übertragen. Dies ist beim PWA1-Verfahren vom Prinzip her die minimale Wortlänge, da sonst die Verbindung zwischen Sender und Empfänger abreißt.

Bei der Übertragung mit einem Bit wird dieses als Vorzeichen interpretiert. Bei der Decodierung wird das Vorzeichen mit der kleinsten Quantisierungsstufe (=1) gewichtet. Auf der Senderseite ist vereinbart, daß in den Pausen der Signalwert 0 durch eine alternierende Folge (+1,-1,+1,...) übertragen wird. Auf der Empfängerseite wird beim Auftreten dieser Folge eine 0 decodiert. Durch diese Maßnahme kann ein störendes Ruhegeräusch vermieden werden.

Nach einer Pause müssen wieder größere Abtastwerte übertragen werden. Es ist eine zusätzliche Maßnahme erforderlich, um den Empfänger schnell auf den ansteigenden Übertragungsbedarf "aufmerksam" zu machen.

Bleibt nach einer alternierenden Folge das Vorzeichen gleich, dann wird eine andere Gewichtung eingeschaltet (=3 im Bereich G11 und =5 in G12), die solange gültig bleibt, bis die Bereiche G11 und G12 verlassen werden.

Geht man von durchgehend gesprochenem Text auf Telefongespräche mit einem Pausenanteil von 60 % über, dann sinkt die mittlere Bitrate auf 2.6 bis 3.2 bit pro Abtastwert (je nach Sprachprobe).

Die subjektive Beurteilung der Simulationsergebnisse ergab, daß abgesehen von einem geringfügig höheren Rauschanteil am Pausenanfang und -ende, kein Unterschied zum Originalsignal feststellbar war.

4.1.5 Extrapolation des Übertragungsbedarfs bei Sprachinterpolation

Unter der Voraussetzung, daß in einem Sprachinterpolationssystem die Quellen voneinander unabhängig sind, und daß alle Quellen die gleiche Verteilung der Wortlängenforderungen haben, kann die Verteilung des momentanen mittleren Bitbedarfs für n Quellen durch $(n-1)$ -fache Faltung der Einzelverteilung berechnet werden (Bild 26).

In Analogie zu der Aussage, daß eine Faltung von Zeitfunktionen einer Multiplikation der entsprechenden Spektren gleichkommt, wurde die $(n-1)$ -fache Faltung der Einzelverteilung durch eine Potenzierung mit n im transformierten Bereich ersetzt. Unter Verwendung der Fast Fourier Transform (FFT), konnte der Rechenaufwand erheblich reduziert werden.

In Bild 27 ist die so berechnete Summenhäufigkeit des mittleren Bitbedarfs in Abhängigkeit von der Quellenzahl n angegeben.

Diese Streuung des momentanen Übertragungsbedarfs erfordert im Überlastungsfall eine Reduzierung einzelner Wortlängen. Es werden bevorzugt solche Bereiche ausgewählt, die im Mittel besser als notwendig übertragen werden. Die Bereiche, deren SNR am nächsten am Toleranzschema liegen, werden im Falle eines Wortlängenüberschusses zuerst mit größerer Wortlänge bedient.

Der Einfluß der Sprachinterpolation durch endliche Quellenzahl ist weitgehend unabhängig vom Codierverfahren. Er wird deshalb exemplarisch beim PWA2-Verfahren (Kapitel 4.2-6) aufgezeigt.

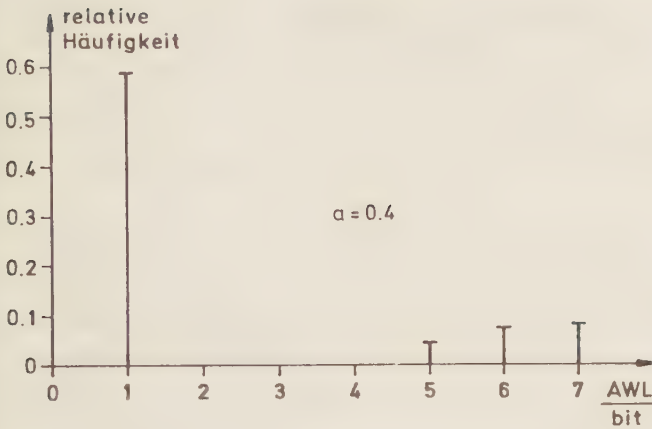


Bild 26: Relative Häufigkeit der angeforderten Wortlängen AWL bei der Einzelquelle (verwendete Sprachproben siehe Abschnitt 5.2)

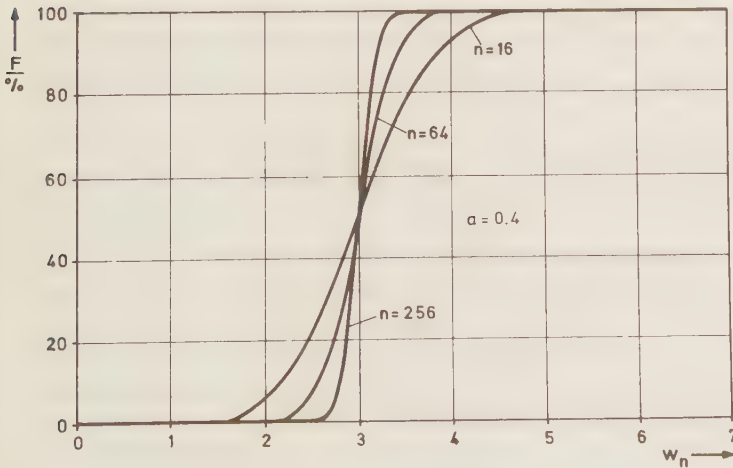


Bild 27: Summenhäufigkeit F des momentanen Bitbedarfs w normiert auf die Quellenzahl n

4.2 PWA2 - Ein verbessertes PWA- Verfahren

Aus Erfahrungen mit der PWA1- Methode ergaben sich Anhaltspunkte für eine Verbesserung.

An erster Stelle ist die hohe Datenrate in den Pausenabschnitten zu nennen.

Zur Aufrechterhaltung des "Gleichlaufs" zwischen Sender und Empfänger muß beim PWA1-Verfahren auch in den Pausen mindestens 1 bit übertragen werden. Dies ist unwirtschaftlich. Wenn man mit einem Sprachdetektor vor der Übertragung abschnittsweise L Pausenabtastwerte von der Übertragung ausschließt, dann genügt es, jeweils für diese L Abtastwerte nur noch 1 Steuerbit zu übertragen.

Diese Maßnahme hat noch einen anderen Vorteil, es kann eine Codierung mit "Pegelregelung" durchgeführt werden. Durch die Trennung zwischen Sprache und Pausen besteht nicht mehr die Gefahr, daß durch die Pegelregelung Pausenabtastwerte verstärkt werden, welche einen unnötigen Übertragungsbedarf erzeugen.

Der Vorteil einer Codierung mit Pegelregelung besteht darin, daß die entstehenden relativen Häufigkeitsverteilungen der geregelten Signale unabhängiger von den Ursprungsverteilungen werden, es tritt eine gewisse Normierung ein /4.2-1, 4.2-2/. Weiterhin erreicht man, daß automatisch ein konstanter SNR-Verlauf über einen weiten Pegelbereich gewährleistet ist.

Eine andere wesentliche Verbesserung betrifft die Übertragung von Daten an Stelle von Sprache. Dieser Aspekt wurde bisher beim PWA1-Verfahren nicht einbezogen.

Ähnlich dem segmentweisen Pausenausschluß kann man mit einem weiteren Steuerbit erzwingen, daß alle Abtastwerte eines Segments mit 8 bit entsprechend der 13-Segmentkennlinie codiert werden. Die Wortlänge von 8 bit wird dann auch durch eine Konkurrenzsituation im Überlastungsfall nicht reduziert. Eine ungestörte Übertragung von Datensignalen ist damit gewährleistet. Selbstverständlich muß auf der Senderseite kontrolliert werden, daß sich nur eine begrenzte Zahl von Datenquellen unter den Sprachquellen befinden. Erreicht dieser Anteil eine bestimmte Grenze, dann müssen weitere Datenquellen zurückge-

wiesen werden, die Leitung ist für sie "belegt".

Mit zwei Steuerbit pro Segment lassen sich 4 Betriebsarten codieren. Bisher wurden 3 davon angesprochen:

Pausen

Daten

Normalbetrieb (DPCM/PCM)

Es hat sich beim PWA1-Verfahren gezeigt, daß die Entscheidung, ob PCM- oder DPCM- Übertragung günstiger ist, aus der Signalvergangenheit nur ungenau geschätzt werden kann. Hier bietet sich mit der 4. Betriebsart eine Steuermöglichkeit an, die in Grenzfällen eine PCM-Übertragung für ein ganzes Segment anordnet. Dabei unterliegt die verwendete Wortlänge der normalen Zuteilung abhängig vom Bedarf der anderen Quellen.

4.2.1 Einfluß des segmentweisen Pausen- ausschlusses auf die Aktivität

Es muß sichergestellt werden, daß durch den abschnittswisen Ausschluß von Abtastwerten keine relevanten Anteile betroffen sind. Deshalb wird vereinbart, daß Segmente, die auch nur einen zu einem Sprachabschnitt gehörenden Abtastwert enthalten, nicht unterdrückt werden dürfen. Am Pausenanfang und -ende werden dadurch im ungünstigsten Fall (L=1) Pausenabtastwerte wie relevante Signale behandelt. Abhängig von der Segmentlänge L erhöht sich scheinbar die Aktivität des Sprachsignals. In Bild 27 ist diese Erhöhung der Aktivität als Funktion von L aufgetragen.

Die Aktivität steigt erst bei großen Segmentlängen merklich an (z.B. bei L=512 um 1%). Für diese Untersuchungen wurden Sprachproben verwendet, die bereits eine Sprach-/Pausenmarkierung durch einen Sprachdetektor enthielten /4.2-3/.

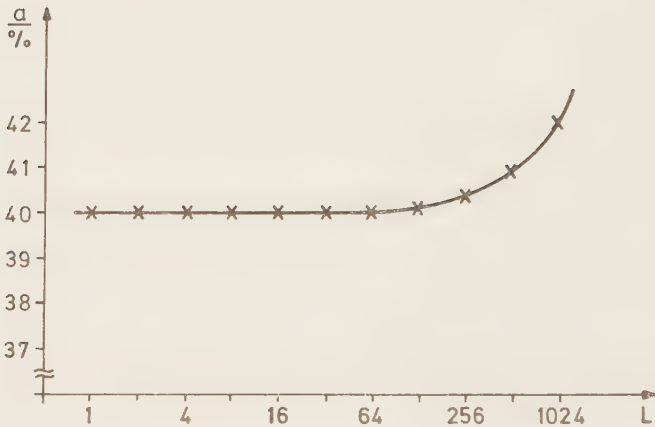


Bild 28: Gemessene Erhöhung der Aktivität a , durch
segmentweisen Pausenausschluß mit der Segmentlänge L

Für die Festlegung der Segmentlänge ist nicht allein die Aktivität maßgebend, sondern auch die damit verbundene Signalverzögerung. Als Kompromiß zwischen Signalisierungsinformation und Verzögerung wird $L=128$ gewählt. Bei 2 Steuerbits (Pausen, Daten, DPCM/PCM, PCM) beträgt die mittlere Bitrate in den Pausen $1/64$ bit pro Abtastwert. Es entsteht eine Signalverzögerung von 16 ms.

4.2.2 Unterscheidung zwischen Normalbetrieb und segmentweiser PCM-Übertragung

Im Normalbetrieb kann, abhängig vom geschätzten Prädiktionsgewinn, jederzeit zwischen PCM- und DPCM-Übertragung umgeschaltet werden.

Bei der segmentweisen PCM-Übertragung wird festgelegt, daß alle Abtastwerte eines Segments mit PCM übertragen werden. Damit lassen sich Grenzfälle ausschalten, wo eine PCM-Übertragung günstiger ist, aber durch die ungenaue Schätzung aus der Signalvergangenheit immer wieder auf eine DPCM-Übertragung umgeschaltet wird.

Das genaueste Entscheidungskriterium ist der gemessene Prädiktionsgewinn für ein Segment von Abtastwerten. Dazu müßten bereits vor der Übertragung PCM- und DPCM-Leistungsmittelwerte bestimmt und daraus der Prädiktionsgewinn berechnet werden. Dies würde eine aufwendigere Systemstruktur bedingen. Deshalb soll ein anderes, einfach zu berechnendes Entscheidungskriterium verwendet werden.

Es hat sich gezeigt, daß die Nulldurchgangsrate f_z gut mit einer "mittleren Frequenz" f_m , die aus dem Kurzzeitspektrum berechnet werden kann, korreliert ist. Geht man von den spektralen Leistungsdichten $S(f_i)$ aus (siehe Bild 29) und bezeichnet man die Gesamtleistung mit S_0 , so soll die mittlere Frequenz f_m (Schwerpunktsfrequenz) durch folgende Gleichung berechnet werden:

$$f_m = \frac{2 \sum_{i=1}^{L/2} f_i \cdot S(f_i)}{S_0}$$

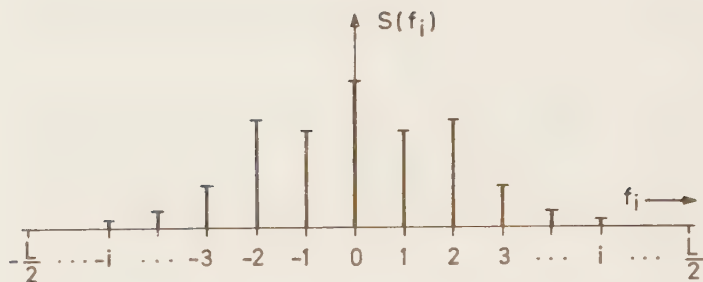


Bild 29: Spektrale Leistungsdichten eines Kurzzeit-spektrums abhängig von der Frequenz f_i

Um die Übereinstimmung zwischen Nulldurchgangsrate f_z und mittlerer Frequenz f_m zu prüfen, wurden beide Größen für Segmente von 128 Abtastwerten bei mehreren Sprachproben gemessen und einer Regressionsanalyse unterworfen. Das Ergebnis in Form der Regressionsgerade ist zusammen mit den gemessenen Werten von f_z und f_m in Bild 30 aufgetragen. Es ist eine sehr gute Übereinstimmung vorhanden.

Für die Festlegung einer Entscheidungsschwelle, ab wann segmentweise PCM übertragen wird, dient die Häufigkeit einer falschen PCM/DPCM-Umschaltung. Signalisiert man bei $f_z = 800\text{Hz}$ eine segmentweise PCM-Übertragung, dann wird, wie eine Messung zeigt, die Zahl falscher PCM/DPCM-Umschaltungen minimal.

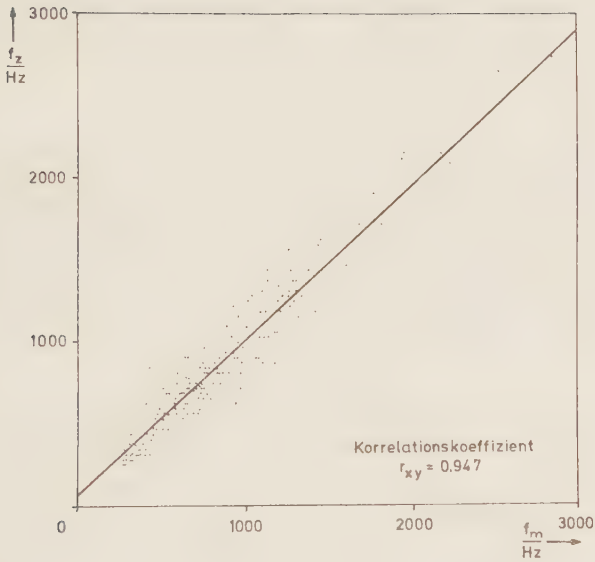


Bild 30: Übereinstimmung von Nulldurchgangsrate f_z und der mittleren Frequenz f_m aufgezeigt durch Regressionsanalyse (durchgezogene Linie: Regressionsgerade, Punkte: Meßwerte, Korrelationskoeffizient: 0.95)

4.2.3 Codierung mit Pegelregelung

Das Prinzip einer Codierung mit Pegelregelung beruht darauf, daß man nicht die Abtastwerte $u(k)$ direkt, sondern die Quotienten $q_s(k)$ quantisiert, die sich durch Division jedes Wertes $u(k)$ durch eine Größe σ_s ergeben:

$$q_s(k) = u(k) / \sigma_s$$

Dabei tritt ein Quantisierungsfehler $\Delta q_s(k)$ auf. Auf der Empfangsseite werden die quantisierten Werte $q_{s1}(k)$ mit σ_s multipliziert und damit die Normierung wieder rückgängig gemacht (siehe Bild 31). Das rekonstruierte Signal wird mit $y(k)$ bezeichnet.

$$q_{s1}(k) = q_s(k) + \Delta q_s(k)$$

$$y(k) = q_{s1}(k) \cdot \sigma_s$$

Der Fehler zwischen Eingangs- und Ausgangssignal berechnet sich zu:

$$\varepsilon_s(k) = [u(k) / \sigma_s + \Delta q_s(k)] \cdot \sigma_s - u(k)$$

$$\varepsilon_s(k) = \Delta q_s(k) \cdot \sigma_s$$

Der Fehler ist also proportional zu σ_s .

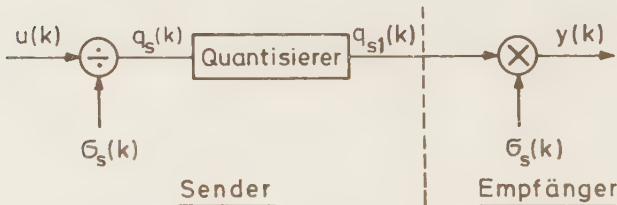


Bild 31: Prinzip der Codierung mit Pegelregelung

Die bekannten Verfahren der Codierung mit Pegelregelung (oder auch adaptive Quantisierung) /4.2-1, 4.2-2/ unterscheiden sich hauptsächlich in der Berechnung von σ_s .

Man spricht von einer Vorwärtssteuerung, wenn $\sigma_s(k)$ für alle L Abtastwerte eines Signalsegments gleich dem Kurzzeiteffektivwert des m -ten Segments σ_m ist.

$$\sigma_m = \sqrt{\frac{1}{L} \sum_{i=m_0}^{m_0+L-1} u^2(i)}$$

($u(m_0)$ = erster Abtastwert im m -ten Segment)

Aus der ursprünglichen Häufigkeitsverteilung der Abtastwerte $u(k)$ (Bild 32) entsteht eine neue Verteilung der Quotienten $q(k)$, die je nach Segmentlänge L unterschiedliche Form hat (Bild 33).

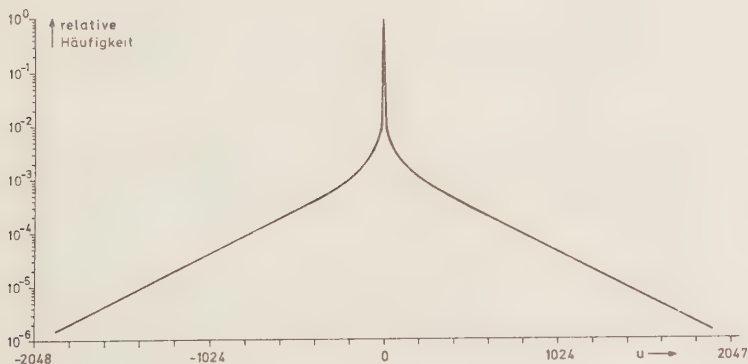


Bild 32: Relative Häufigkeit von Signalspannungen $u(k)$ bei Sprache
(verwendete Sprachproben siehe Abschnitt 5.2)

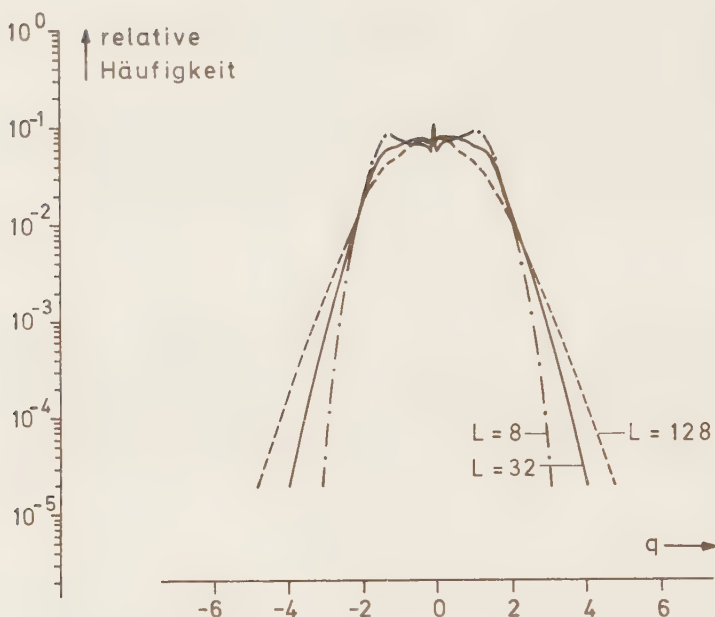


Bild 33: Relative Häufigkeit der pegelgeregelten Signale q bei Vorwärtssteuerung mit der Segmentlänge L (Sprachproben siehe 5.2)

Damit wieder mit ganzen Zahlen gearbeitet werden kann, muß noch eine Bereichstransformation mit einem konstanten Faktor durchgeführt werden (z.B. Abbildungen von q auf das Intervall $-2048 \dots 2047$).

Da bei der Vorwärtssteuerung σ_m bereits vor der Übertragung von Signalabtastwerten auch dem Empfänger bekannt sein muß, bleibt nur noch übrig, σ_m selbst vorher zum Empfänger zu übertragen. Diese Möglichkeit soll beim PWA-Verfahren ausscheiden. Es wird das Prinzip der Rückwärtssteuerung angewendet. Dabei ist für jeden Quotient q ein spezielles σ_s aus der Signalvergangenheit zu berechnen.

$$\sigma_s(k) = \sqrt{\frac{1}{L} \sum_{i=1}^L u^2(k-i)}$$

Diese Schätzung des Kurzzeiteffektivwerts aus der Signalvergangenheit kann auch auf der Empfangsseite vorgenommen werden. Es ergeben sich wieder verschiedene Verteilungen der Quotienten abhängig von der Segmentlänge L (Bild 34).

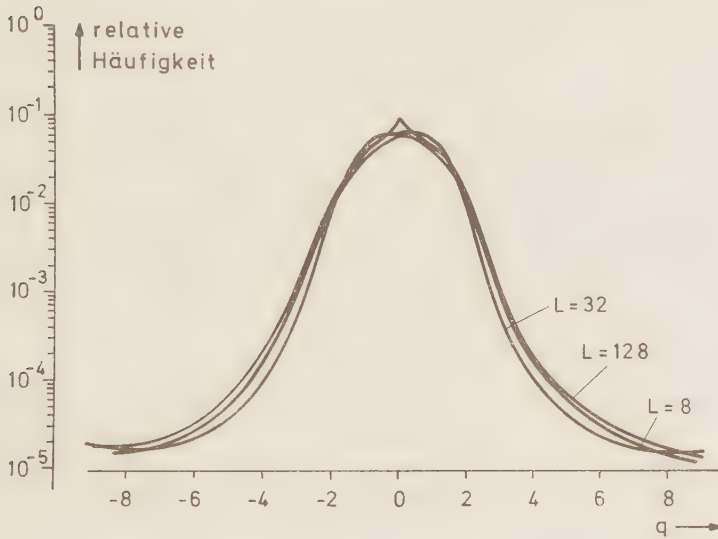


Bild 34: Relative Häufigkeit der pegelgeregelten Signale q bei Rückwärtssteuerung mit der Segmentlänge L (Sprachproben siehe 5.2)

Aus der Form der Verteilung kann bereits die Problematik der Schätzung des Signaleffektivwerts aus der Signalvergangenheit abgelesen werden. Es treten im Vergleich zur Vorwärtssteuerung sehr viel größere Werte von q auf. Die Auftretenswahrscheinlichkeit ist zwar gering, aber man muß diese Übersteuerungsgefahr berücksichtigen. Ähnliche Überlegungen wie beim PWA1-Verfahren zeigen, daß hier ein Kompromiß bezüglich der Segmentlänge L getroffen werden muß. Bei den Untersuchungen wurde $L = 8$ als günstigster Wert ermittelt.

Der Einfluß der Codierung mit Pegelregelung auf den Störab-

stand kann am besten an einigen Beispielen verdeutlicht werden. In Bild 35 ist unter Verwendung einer linearen Kennlinie mit 128 Stufen (7 bit) der Unterschied zwischen einer Codierung mit- und ohne Pegelregelung dargestellt.

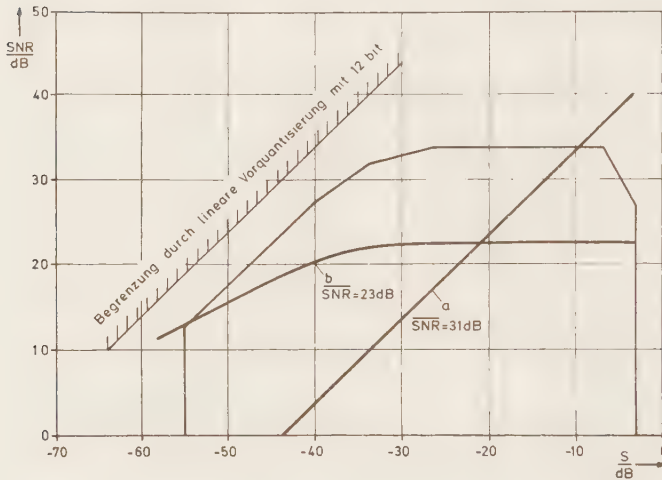


Bild 35: Unterschied zwischen einer Codierung (a) ohne und (b) mit Pegelregelung. Quantisierungskennlinie: 7 bit linear (Sprachproben siehe 5.2)

Im nächsten Bild (Bild 36) wurde eine nichtlineare Quantisierungskennlinie benutzt. Es wurde dabei eine logarithmische Kennlinie durch eine 13-Segment-Kennlinie mit 7 bit angenähert. In diesem Bild sind auch Ergebnisse aufgetragen, die bei Verwendung von DPCM-Werten an Stelle der PCM-Werte bei der Pegelregelung entstehen. Die Quotienten $q_d(k)$ werden ganz ähnlich berechnet:

$$q_d(k) = d(k) / \sigma_s(k)$$

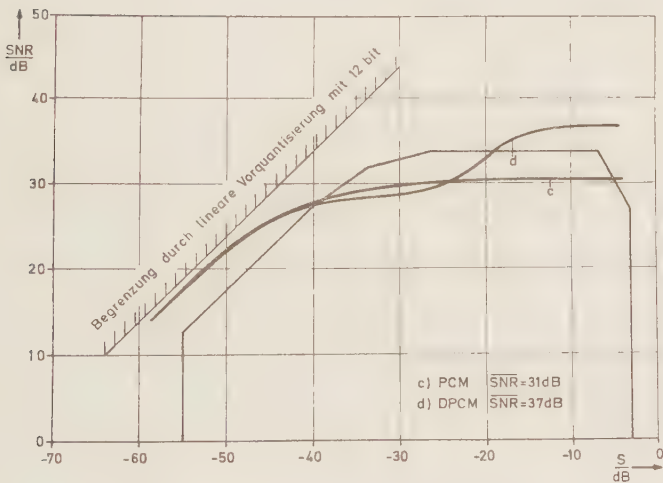
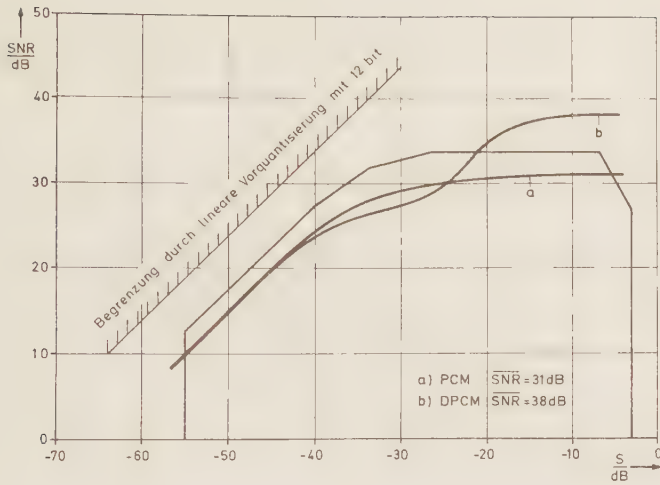


Bild 36: Quantisierung durch 13-Segmentkennlinie (7 bit)
 (a, b) ohne und (c, d) mit Pegelregelung
 (Sprachproben siehe Abschnitt 5.2)

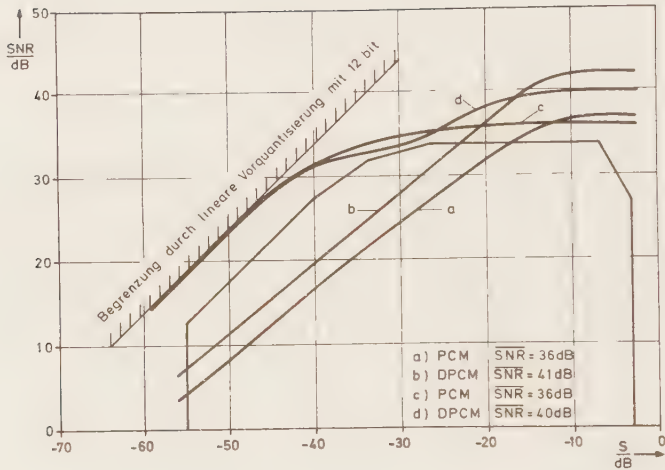


Bild 37: Quantisierung durch optimale Kennlinie (7 bit)
(a, b) ohne und (c, d) mit Pegelregelung

Schließlich wurden noch aus den entsprechenden Häufigkeitsverteilungen optimale Kennlinien berechnet. Die Ergebnisse sind für Wortlängen von 7 bit in Bild 37 zusammengestellt.

Ein Vergleich der Kurven c) und d) in Bild 37 zeigt, daß bei einer Codierung mit Pegelregelung bereichsweise mit einer PCM-Übertragung bessere Ergebnisse erzielt werden können als mit DPCM-Übertragung.

4.2.4 Signalklassifizierung bei Codierung mit Pegelregelung

Beim PWA2-Verfahren ist durch die automatische Anpassung des Störabstandes über einem weiten Signalpegel eine Klassifizierung nach dem Signalpegel nicht mehr sinnvoll.

Abhängig vom geschätzten Prädiktionsgewinn G_p wird eine Unterscheidung der pegelgeregelten Signale nach DPCM- und PCM-Übertragung vorgenommen. Bei Prädiktionsgewinnen $G_p > 0$ soll mit DPCM, sonst mit PCM übertragen werden.

Es hat sich gezeigt, daß eine weitere Differenzierung des PCM-Bereichs keine Vorteile bringt. Der DPCM-Bereich dagegen kann entsprechend des Prädiktionsgewinns in 6 Teilbereiche aufgespalten werden. Eine zusätzliche Verbesserung erhält man durch weitere Aufteilung dieser 6 Bereiche nach der Tendenz des Prädiktionsgewinns (ansteigender - abfallender Gewinn).

Damit ist eine Unterscheidung in 13 Signalbereiche getroffen worden. 12 Bereiche sind für DPCM-Übertragung (Nr. 1...6 für ansteigenden Prädiktionsgewinn, Nr. 7...12 für fallenden Gewinn) vorgesehen und ein Bereich (Nr. 13) ist für PCM-Übertragung reserviert.

In jedem Bereich wurden wie beim PWA1-Verfahren Häufigkeitsverteilungen gemessen. Dabei hat sich gezeigt, daß durch Normierungsfaktoren C_y einige Verteilungen in einander überführt werden können.

Schließlich sind nur noch 7 Quantisierungskennlinien (KL1-KL7) für die 13 Bereiche erforderlich.

In Tabelle 2 ist die Klassifizierung der pegelgeregelten Signale mit den verwendeten Kennlinien und den Normierungsfaktoren C_y zusammengestellt.

Tendenz	Signal- bereich γ	Grenzen für G_p /dB	Normierungs- konstante C_γ	Kennlinien- Nr.
DPCM steigender Prädiktions- Gewinn	1	>15	5.6	KL1
	2	12 ... 15	4.8	KL2
	3	9 ... 12	3.6	KL3
	4	6 ... 9	2.8	KL3
	5	3 ... 6	2.0	KL4
	6	0 ... 3	1.5	KL4
DPCM fallender fallender Prädiktions- Gewinn	7	>15	4.8	KL5
	8	12 ... 15	4.0	KL5
	9	9 ... 12	3.0	KL5
	10	6 ... 9	2.0	KL5
	11	3 ... 6	1.4	KL6
	12	0 ... 3	1.0	KL6
PCM	13	<0	1.0	KL7

Tabelle 2: Klassifizierung der pegelgeregelten Signale

4.2.5 Codierung der Einzelquelle

Die Optimierung der 7 Kennlinien aus den entsprechenden Häufigkeitsverteilungen erfolgt so, wie im Abschnitt 4.1.3 beschrieben.

Je nach verwendeter Wortlänge kann für jeden Teilbereich ein bestimmter Störabstand berechnet werden. Die Forderung nach einem konstanten Störabstand von 34 db legt in jedem Bereich die zu verwendende Wortlänge fest (Tabelle 3).

Tendenz	Signalbereich	Kennlinien-Nr.	verwendete Wortlängen in bit	mittlerer Störabstand SNR in dB
DPCM steigender Prädiktions- Gewinn	1	KL1	6	33.0
	2	KL2	6	35.3
	3	KL3	6	37.4
	4	KL3	6	37.5
	5	KL4	6	34.6
	6	KL4	7	36.7
DPCM fallender Prädiktions- Gewinn	7	KL5	6	33.4
	8	KL5	6	35.9
	9	KL5	5	32.4
	10	KL5	5	32.0
	11	KL6	6	37.1
	12	KL6	6	35.3
PCM	13	KL7	7	36.0

Tabelle 3: Verwendete Wortlängen und damit erzielte mittlere Störabstände

Teilt man einer Quelle die in Tabelle 3 angegebenen Wortlängen in den entsprechenden Gruppen zunächst ohne Rücksicht auf die Verfügbarkeit zu, dann erhält man als Simulationsergebnis den SNR-Verlauf über dem Signalpegel in Bild 38. Die SNR-Mittelwerte liegen über dem Toleranzschema. In Bild 39 ist zur Charakterisierung der Streuung die Summenverteilung der Segment-SNR-Abstände zum Toleranzschema aufgetragen.

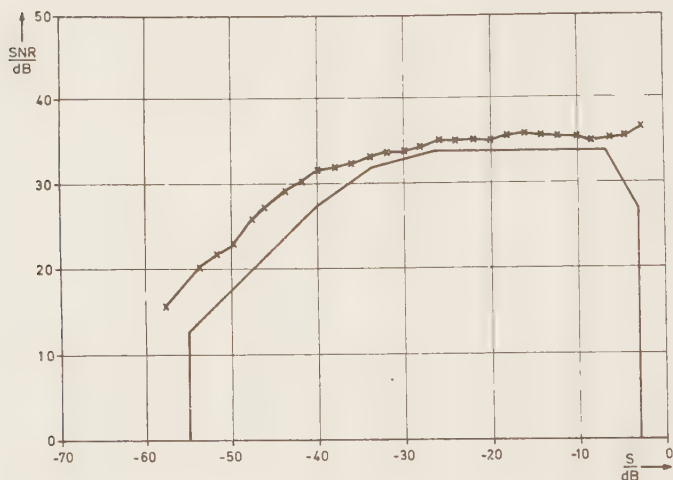


Bild 38: Verlauf des Störabstandes SNR über dem Signalpegel bei Codierung einer Einzelquelle ohne Sprachinterpolation

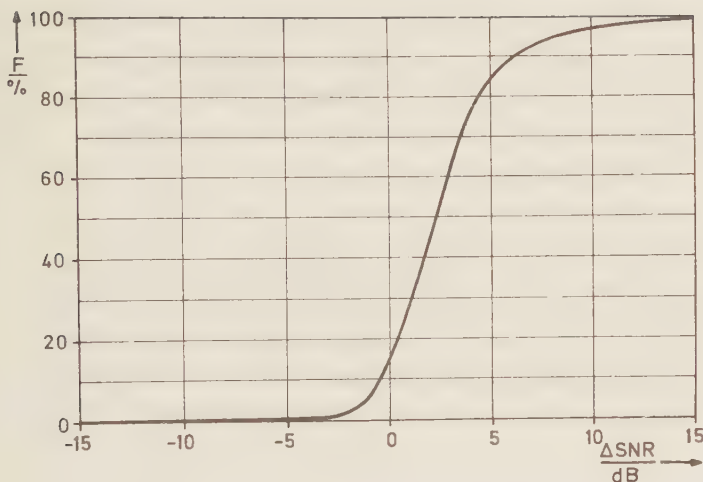


Bild 39: Summenhäufigkeit des SNR-Abstandes zum Toleranzschema bei Codierung einer Einzelquelle ohne Sprachinterpolation

Wie man daraus ablesen kann, liegen nur 15 % aller Segment-SNR-Werte unter dem Toleranzschema und nur 1 % liegen weiter als 3 dB darunter. Der lineare Mittelwert der Segment-SNR-Abstände zum Toleranzschema beträgt 2.7 dB. Die 1- σ -Streuengrenze beträgt 3.2 dB. Im Mittel müssen bei dieser Wortlängenzuteilung für durchgehend gesprochenen Text 5.6 bit pro Abtastwert, für Telefonsprache mit 40 % Aktivität 2.7 bit pro Abtastwert aufgewendet werden.

Bei dieser Bilanz ist enthalten, daß durch die segmentweise PCM-Übertragung auch jeweils 128 Abtastwerte mit 7 bit übertragen werden. Ebenfalls enthalten ist die Steuerinformation von 1/64 bit pro Abtastwert, mit der die vier Betriebsarten (Pausen, Daten, usw.) selektiert werden können.

In Tabelle 4 ist die relative Häufigkeit, mit der die verwendeten Wortlängen auftreten, angegeben. Die Steuerinformation

von $1/64$ bit wird hier näherungsweise mit Null angesetzt.

Wortlänge (bit)	relative Häufigkeit (Sprache mit $a=40\%$)
0	59.6 %
5	5.8 %
6	21.2 %
7	13.4 %

Tabelle 4: Auftrittshäufigkeit der verwendeten Wortlängen

Geht man wie beim PWA1-Verfahren davon aus, daß alle Quellen etwa die gleiche Verteilung der Wortlängenforderungen haben, und daß die Quellen voneinander statistisch unabhängig sind, so erhält man die Verteilung des mittleren Wortlängenbedarfs für n Quellen durch $(n-1)$ -fache Faltung der Einzelverteilung (Tabelle 4).

In Bild 40 ist die auf n normierte Summenhäufigkeit der Wortlängenforderungen dargestellt.

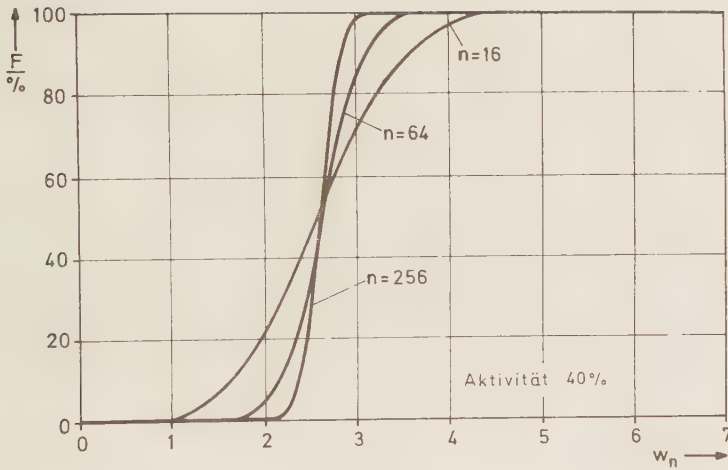


Bild 40: Summenhäufigkeit F der Wortlängen-
anforderungen W normiert auf die
Quellenzahl n

4.2.6 PWA2 - Simulation mit Scharmittlung über eine endliche Quellenzahl

Während bei der Codierung der Einzelquelle ohne Sprachinterpolation eine Quelle die angeforderte Wortlänge auch erhält, muß bei der Scharmittlung eine vorgegebene feste Länge des Übertragungsrahmens eingehalten werden. Je nach Quellenzahl streut der momentane Übertragungsbedarf mehr oder weniger um den Mittelwert. Ist die Summe der angeforderten Wortlängen gleich der Länge des Übertragungsrahmens, dann sind keine weiteren Maßnahmen notwendig. Im Überlastungsfall muß ein Teil der angeforderten Wortlängen reduziert werden. Wird die vorhandene Übertragungskapazität durch die Anforderungen nicht ausgeschöpft, dann kann einzelnen Quellen eine größere Wortlänge zugeteilt werden. Die Wortlängenreduktion bzw. -erweiterung erfolgt nach einer vorgegebenen Prioritätsstrategie.

Zur Festlegung einer Prioritätsliste muß bekannt sein, welcher Quellenzustand am leichtesten eine Wortlängenreduktion zuläßt oder für welchen Zustand eine Wortlängenerweiterung am notwendigsten ist. Dazu kann man sich der in Tabelle 3 angegebenen Störabstände bedienen. Höchste Priorität erhält der Signalzustand, dessen SNR mit der vereinbarten Wortlänge am weitesten unter dem angestrebten Wert von 34 dB liegt. Die niedrigste Priorität dagegen hat der Quellenzustand, dessen Störabstand am weitesten über 34 dB liegt.

Im Überlastungsfall wird eine Wortlängenreduktion um 1 bit, beginnend bei der Quelle durchgeführt, deren Zustand die niedrigste Priorität aufweist. Dies wird solange in Richtung höherer Prioritäten fortgesetzt, bis die den Quellen zugeteilten Wortlängen mit der Übertragungskapazität übereinstimmt (notfalls in mehreren Durchläufen).

Umgekehrt wird bei einer Wortlängenerweiterung die Quelle mit der höchsten Priorität zuerst mit einem zusätzlichen Bit bedacht u.s.w.

Wenn mehrere Quellen die gleiche Priorität anzeigen, wird durch eine zyklisch veränderliche Reihenfolge sichergestellt, daß im Mittel keine Quelle bevorzugt oder benachteiligt wird.

Die Ergebnisse einer Simulation des PWA2 Verfahrens mit Scharmittelung über 128 Quellen sind für Datenraten zwischen 16 kbit/s bis 24 kbit/s in Form des mittleren SNR-Verlaufs über dem Signalpegel in Bild 41 aufgetragen. Die dazugehörigen Verteilungen der SNR-Abstände zum Toleranzschema sind in Bild 42 als Summenverteilungen dargestellt. Die Abhängigkeit der SNR-Streuung um das Toleranzschema von der Zahl der Quellen geht aus Bild 43 hervor. Dort sind für die Datenrate von z.B. 24 kbit/s die SNR-Abstände zum Toleranzschema, abhängig von der Quellenzahl, als Summenverteilungen dargestellt.

Damit der mittlere Störabstand über dem Toleranzschema liegt, ist bei Scharmittelung über 128 Quellen eine Datenrate von 22 kbit/s notwendig. Durch subjektive Beurteilung (Hörvergleich) ist auch bei 20 kbit/s noch keine hörbare Störung wahrnehmbar. Erst bei 16 kbit/s und niedriger Quellenzahl z.B. 16 ist eine geringe Störung feststellbar.

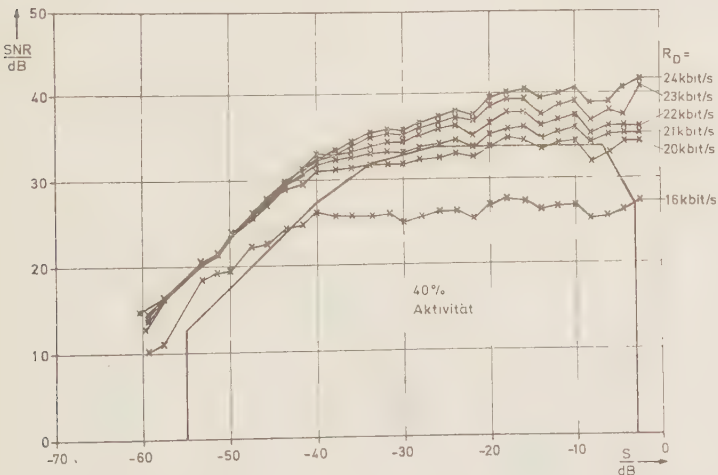


Bild 41: Verlauf des Störabstandes SNR über dem Signalpegel bei Sprachinterpolation mit 128 Quellen und verschiedenen Datenraten R

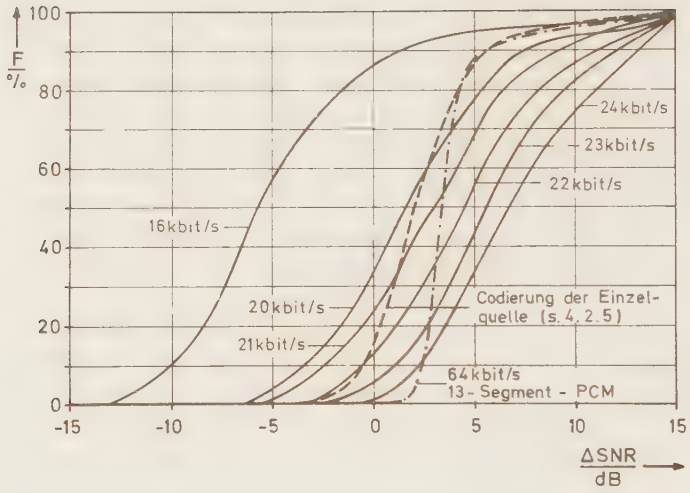


Bild 42: Summenhäufigkeit des SNR-Abstandes zum Toleranzschema bei Sprachinterpolation mit 128 Quellen und verschiedenen Datenraten

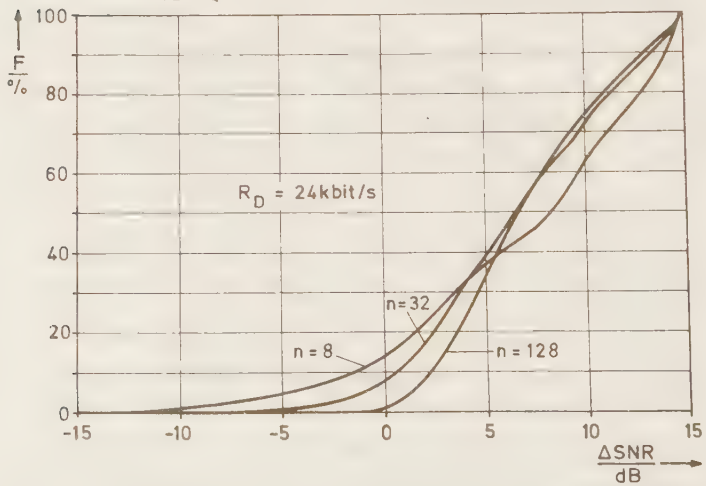


Bild 43: Summenhäufigkeit des SNR-Abstandes zum Toleranzschema bei einer Datenrate von 24 kbit/s und verschiedenen Quellenzahlen bei der Sprachinterpolation

Da für alle Signalpfade das gleiche Blockschaltbild gilt, soll die Erklärung der Funktionen an einem beliebig herausgegriffenen Pfad erfolgen.

Sowohl auf Sender- wie auch auf Empfängerseite ist jeweils ein vollständiges DPCM- und PCM-System vorhanden. Die Entscheidung, ob codierte DPCM-Werte oder codierte PCM-Werte übertragen werden, wird von einer Klassifizierungslogik gefällt.

Als Entscheidungskriterium wird aus der Signalvergangenheit der Wert $S^*(k)$ und $D^*(k)$ (Kurzzeiteffektivwert der PCM- und der DPCM-Werte) berechnet. Damit ist vereinbarungsgemäß eine Klassifizierung in 9 Gruppen möglich.

Für jede Gruppe steht in einer Tabelle die notwendige Wortlänge und die Nummer der zur Gruppe gehörenden Quantisierungskennlinie zur Verfügung.

Die notwendige Wortlänge wird von allen Signalpfaden als Wortlängenanforderung an die Zentralsteuerung gemeldet. Dort wird die Summe dieser Anforderungen mit der Kanalkapazität verglichen. Bei Übereinstimmung wird jedem Signalpegel die angeforderte Wortlänge zugeteilt.

Sind die Anforderungen zu hoch, so werden reduzierte Wortlängen zugeteilt. Im anderen Fall kann eine größere Wortlänge als angefordert wurde, zugeteilt werden. Die Reduzierung oder Erweiterung der Wortlängen erfolgt nach Prioritätsregel. Zu jeder Signalgruppe gehört eine Prioritätsziffer.

Ist eine Wortlängenreduktion erforderlich, dann werden zunächst die zu der niedrigsten Priorität gehörenden Wortlängen um 1 bit gekürzt.

Dies wird solange in Richtung steigender Priorität fortgesetzt, bis die Summe der zugeteilten Wortlängen mit der Kanalkapazität übereinstimmt.

Die minimale Wortlänge für jede Quelle beträgt (auch in den Pausen) 1 bit.

Bei Wortlängenerweiterung wird zuerst die Quelle mit der höchsten Priorität mit einem zusätzlichen Bit bedient, usw.

Mit der zugeteilten Wortlänge kann in jedem Signalpfad die Codierung ausgeführt werden. Die unterschiedlich langen Codeworte werden zu einem Rahmen mit konstanter Länge zusammengefügt.

Im Abschnitt 4.4 sind dazu noch zusätzliche Anmerkungen gemacht.

Auf der Empfangsseite wird für jeden Signalpfad die zugeteilte Wortlänge und die entsprechende Signalgruppe genauso wie auf der Senderseite bestimmt. Damit können aus dem Übertragungsrahmen die individuellen Codeworte entnommen und nach einer DPCM/PCM-Entscheidung decodiert werden.

4.3.2 PWA2-System

Bei diesem Verfahren wird durch eine übergeordnete Steuerung zunächst für jeweils z.B. 128 Abtastwerte einer Quelle entschieden, ob es sich um Pausenabtastwerte, Sprache oder Daten handelt. Dabei wurde vorausgesetzt, daß der dazu notwendige Sprachdetektor bereits vorhanden ist /4.2-3/. Das gleiche gilt für den Datendetektor.

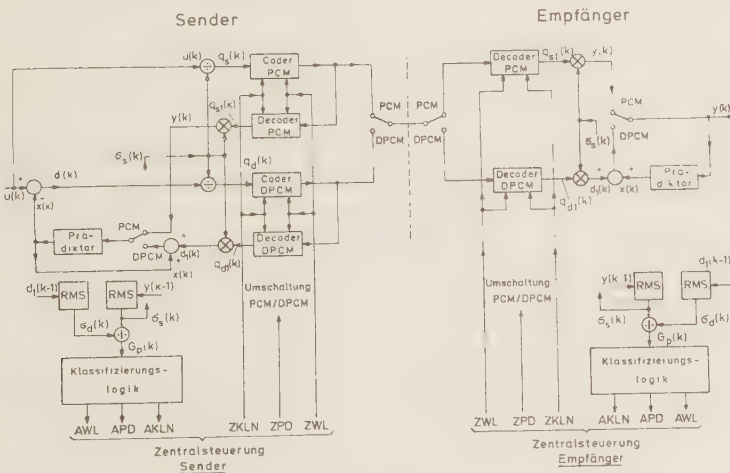


Bild 45: Blockschaltbild eines Signalpfades im PWA2-System

angeforderte: A $\begin{matrix} \diagup \\ \diagdown \end{matrix}$ WL Wortlänge
 Z $\begin{matrix} \diagup \\ \diagdown \end{matrix}$ PD PCM/DPCM-Umschaltung
 zugeteilte: Z $\begin{matrix} \diagup \\ \diagdown \end{matrix}$ KLN Kennlinien-Nr.

Wenn es sich um Sprache handelt, kann noch vereinbart werden, daß ausschließlich mit PCM übertragen werden soll. Dies steht im Gegensatz zum Normalfall, wo nach jedem Sprachabtastwert erneut eine DPCM/PCM- Umschaltung möglich ist. Diese Entscheidungen werden durch zwei Signalisierbits (pro Quelle und 128 Abtastwerte) den Empfängern mitgeteilt. Damit ist es möglich, Pausenabtastwerte von der Übertragung auszuschließen, oder z.B. bei Daten der Quelle eine feste unveränderliche Wortlänge von 8 bit/Abtastwert zuzuordnen. Der wesentlichste Unterschied zum PWA1-System ist durch die Codierung mit Pegelregelung gegeben, die erst durch den Pausenausschluß sinnvoll wird. Dabei werden die PCM- und die DPCM- Werte vor der Quantisierung im Coder durch den Wert $\sigma_s(k)$ dividiert. σ_s ist der Kurzzeiteffektivwert von 8 zurückliegenden PCM-Werten. Es muß dafür gesorgt werden, daß der Fall $\sigma_s = 0$ nicht auftreten kann (z.B. durch Ersetzen: wenn $\sigma_s = 0$, dann wird $\sigma_s = 1$ gesetzt). Deshalb ist eine Klassifizierung nach dem Signalpegel nicht mehr notwendig.

Die Klassifizierung wird nach dem Kurzzeit-Prädiktionsgewinn vorgenommen. Ganz ähnlich wie beim PWA1-System ist auch hier für jeden klassifizierten Signalbereich die notwendige Wortlänge und die Prioritätsziffer in einer Tabelle abgelegt. Dabei kann auch die Wortlänge 0 bit für Pausenabtastwerte auftreten.

Aus den Wortlängenanforderungen wird, ähnlich wie beim PWA1-System, die zugeteilte Wortlänge berechnet. Dabei werden allerdings die von der übergeordneten Steuerung festgelegten Entscheidungen (Daten, Pausen, usw.) berücksichtigt.

Codierung, Decodierung und Umschaltung zwischen DPCM und PCM wurden aus dem PWA1-System übernommen. Aus diesem Grund erhält man für einen Signalpfad im PWA2-System ein, dem Bild 44 sehr ähnliches Blockschaltbild, das sich im wesentlichen nur durch eine zusätzliche Division durch σ_s vor der Codierung und einer Multiplikation mit σ_s nach der Decodierung unterscheidet (Bild 45).

4.4 Maßnahmen bei Fehlern auf dem Übertragungskanal

Bei den bisherigen Ausführungen wurde ein störungsfreier Übertragungskanal vorausgesetzt. Obwohl der Einfluß von Übertragungsfehlern nicht Gegenstand dieser Arbeit ist, sollen einige Anmerkungen zu diesem Themenkreis gemacht werden.

Die Berechnung der Rahmenaufteilung aus der Signalvergangenheit ist bei Fehlern auf dem Übertragungskanal nicht unproblematisch. Wird z.B. auf der Empfängerseite den Quellen andere Wortlängen als auf der Senderseite zugeordnet, dann kann die Verbindung zusammenbrechen, wenn man die zu den einzelnen Quellen gehörenden Worte einfach aneinander reiht (Bild 44).

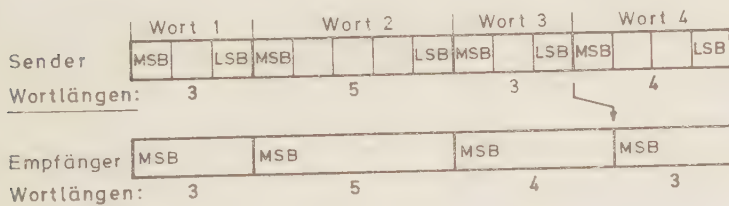


Bild 44: Störanfällige Einteilung des Übertragungsrahmens

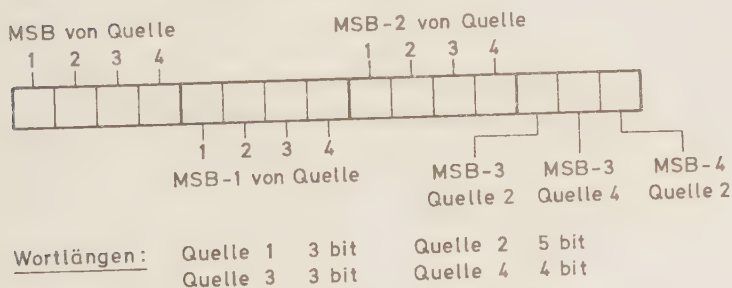


Bild 45: Störnunempfindlichere Einteilung des Übertragungsrahmens

Ordnet man dagegen die einzelnen Bits der verschiedenen Worte anders im Übertragungsrahmen an, dann wird diese Gefahr wesentlich geringer (Bild 45)

Wie man leicht einsieht, wirkt sich eine falsche Wortlängenaufteilung auf der Empfangsseite nicht mehr auf die MSBs der einzelnen Worte direkt aus, sondern nur noch auf niederwertige Bitstellen. Der dadurch entstehende Fehler ist wesentlich geringer. Dies gilt um so mehr, je näher die Einteilung an eine Gleichaufteilung kommt.

Selbstverständlich muß der ungünstigste Fall, der Systemzusammenbruch, berücksichtigt werden. Es ist sicherzustellen, daß nach kürzester Zeit wieder eine Synchronisierung zwischen Sender und Empfänger statt findet. Auch hier wirkt sich die segmentweise Verarbeitung beim PWA2-Verfahren günstig aus, da vereinbart werden kann, daß zu Beginn jedes Segmentes (L-128) die Wortlängen der beteiligten Quellen gleich lang sein sollen. Schließlich ist noch zu erwähnen, daß nach jeder Pause die Übertragung mit einem definierten Anfangszustand beginnt, so daß spätestens bei Beginn einer Pause Fehler "vergessen" werden. Sind an der Übertragung Datenquellen beteiligt, dann ist es kein Problem, die Datenworte unabhängig von der übrigen Wortaufteilung richtig zu übertragen. Es muß nur vereinbart werden, daß alle Datenquellen zusammengefaßt am Anfang des Übertragungsrahmens mit der festen Wortlänge von 8 bit pro Datenwort übertragen werden. Da die Zahl der Datenquellen aus der Signalisierungsinformation für jedes Segment entnommen werden kann, ist es einfach, den Rest des Übertragungsrahmens, wie besprochen, an die übrigen Quellen aufzuteilen.

5. Computersimulation

In diesem Abschnitt sollen einige, bei den Untersuchungen gültige Randbedingungen angesprochen werden.

Grundlage für die Bereitstellung von Sprachsignalen bei der Computersimulation ist die in 5.1 beschriebene Anordnung zur Ein-/Ausgabe von Sprachsignalen am Institutsrechner. In 5.2 sind einige Angaben über die verwendeten Sprachproben gemacht. Im Abschnitt 5.3 wird auf die Programmierung der Systemsimulation eingegangen.

5.1 Sprachsignal-Ein/Ausgabe am Institutsrechner HP2100

Bei der Simulation von sprachverarbeitenden Systemen kann auf ein Interface für Analogsignale nicht verzichtet werden. Es ist notwendig, als Eingangssignal reale Sprache zu verwenden.

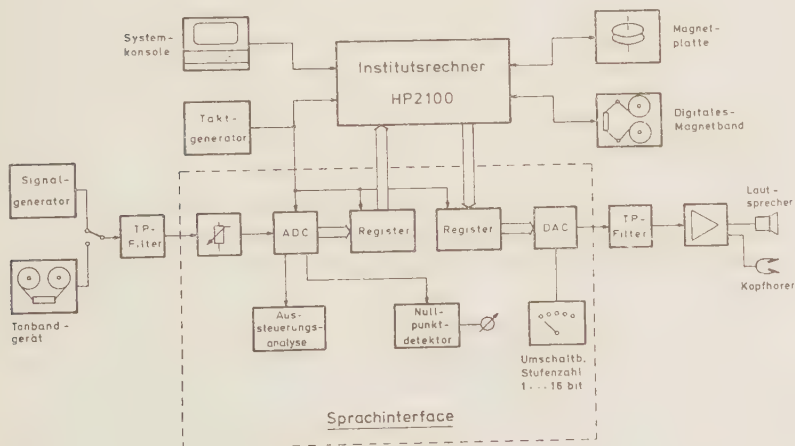


Bild 48. Blockschaltbild der Sprachsignalein-/ausgabe

Außerdem muß die Möglichkeit vorhanden sein, das Simulationsergebnis subjektiv zu beurteilen. Dies ist deshalb sehr wichtig, weil am Ende des Nachrichtenübertragungssystems das menschliche Gehör als Sinke steht. Schließlich ist auch ein objektives Qualitätskriterium, wie z.B. der Störabstand nur dann als absoluter Maßstab sinnvoll, wenn eine Relation zum subjektiven Höreindruck vorhanden ist. In Bild 48 ist die gerätechtechnische Anordnung des Sprachsignalinterfaces als Blockschaltbild wiedergegeben.

Im Systemmittelpunkt steht der Institutsrechner (HP 2100), ein Kleinrechner der Firma Hewlett Packard. Neben der üblichen Standardperipherie interessiert hier nur die Systemkonsole, über die im Dialog die notwendigen Steueranweisungen eingegeben werden, und die Hintergrundspeicher in Form einer Magnetbandstation und einer Magnetplatteneinheit.

Für die Analog-Digital-Umsetzung wurde ein hochwertiger 12 bit Wandler (FHN 5002) der Firma Friesseke und Hoepfner an den Rechner angepaßt. Der Wandler zeichnet sich insbesondere durch eine sehr geringe differentielle Nichtlinearität aus ($< 1\%$), die für Amplitudendichteverteilungsmessung an Sprachsignalen von großer Bedeutung ist. Auf der Ausgangsseite ist ein 16 bit-D/A-Wandler (Hybrid Systems, DAC 355-16) vorhanden, mit dem die volle Wortlänge des Rechners ausgenutzt werden kann. Durch Umschaltung der Wortlänge können 1 ... 16 bit-Wandler simuliert werden.

Bei Untersuchungen an steil abfallenden Amplitudendichteverteilungen ist eine genaue Justierung des Nullpunkts wichtig. Durch eine sehr einfache Schaltung kann dabei der Nullpunkt auf den Bruchteil einer Quantisierungsstufe eingestellt werden. Aus dem digitalen Ausgangssignal des A/D-Wandlers werden die 3 Zustände kleiner Null, gleich Null und größer Null dekodiert und mit einem Zeigerinstrument (Nullpunkt in der Mitte) angezeigt. Der Abgleich erfolgt dann mit einem sehr kleinen Eingangssignal. Durch die Trägheit des Zeigers wird ein Mittelwert entsprechend der Häufigkeit von -1, 0 und +1 angezeigt, der bei richtigem Abgleich Null werden muß.

Ein weiteres Problem stellt die reproduzierbare Aussteuerung

durch das Eingangssignal dar. Die Messung des mittleren Effektivwerts läßt sich digital nur mit größerem Aufwand durchführen. Sehr viel einfacher kann die Aussteuerung durch Messung der Überschreitungswahrscheinlichkeit einer bestimmten Signalamplitude (durch Zählen) kontrolliert werden. Dieses Verfahren ist hier zweckmäßig, da der A/D-Wandler entsprechende Ausgänge aufweist.

Die Ein- und Ausgabe der Signalabtastwerte muß synchron mit der Abtastfrequenz erfolgen. Beträgt diese bei Sprache z.B. 8 kHz, so muß bei einer Echtzeitverarbeitung neben der Ein- und Ausgabe auch noch die Verarbeitung in 125 µs durchgeführt werden. Dies ist nur für sehr einfache Verarbeitungsalgorithmen (z.B. DPCM) möglich.

Bei der Simulation von komplexeren Systemen muß ein anderer Weg eingeschlagen werden. Zunächst werden die Sprachsignale in Echtzeit auf einem Hintergrundspeicher (Platte, Band) abgelegt. Die eigentliche Simulation kann dann ohne Echtzeitbedingungen ablaufen. Die Simulationsergebnisse werden ebenfalls auf den Hintergrundspeicher gebracht und in einem 3. Schritt wieder in Echtzeit über das Sprachinterface ausgegeben.

Im Gegensatz zur synchronen Ein- oder Ausgabe der Abtastwerte, erfolgt der Datentransfer mit Hintergrundspeichern immer blockweise, also asynchron. Dabei muß noch berücksichtigt werden, daß dem Datentransfer eine Wartezeit vorausgeht, die durch das Positionieren der Schreib/Leseköpfe bei einer Magnetplatte, oder durch das Beschleunigen der Antriebsmechanik bei einer Magnetbandstation bedingt ist. Die Kopplung zwischen dem synchronen und dem blockweisen Transfer kann mit dem Wechselfufferprinzip vorgenommen werden. Dabei werden zwei Datenfelder abwechselnd synchron geladen und blockweise ausgegeben oder umgekehrt.

Die besonderen zeitlichen Anforderungen bei den Ein-/ Ausgabe-Operationen können nur unter Umgehung des Betriebssystems des Rechners erfüllt werden. Dazu ist an den zeitkritischen Stellen eine Programmierung auf Assemblerebene erforderlich. Bei dem Wechselfufferverfahren wird ein DMA-Betrieb (direct memory access) über 2 unabhängige Kanäle gleichzeitig durchgeführt. Daneben wurde im Interesse einer einfachen Bedienbar-

keit ein ausführlicher Dialog programmiert, der dem Benutzer Erklärungen zu den erforderlichen Steuerkomandos gibt. Komplexe Steuerungsabläufe, wie z.B. die Bedienung eines (fernsteuerbaren) Tonbandgerätes konnten dabei automatisiert werden. So wurde es möglich, auf verschiedenen Spuren des Tonbandes Originalsprache und Simulationsergebnisse mit exakter zeitlicher Zuordnung rechnergesteuert auszugeben, so daß ein direkter Vergleich bei der Wiedergabe durch Umschalten der Spuren möglich ist.

Die realisierte Sprachsignalein-/ausgabe kann analoge Signale mit bis zu 33 kHz abtasten und diese mit 12 bit quantisierten Abtastwerte in Echtzeit auf einer Magnetplatte oder einer Magnetbandstation ablegen. Mit der gleichen Rate können Abtastwerte mit 16 bit Wortlänge auch wieder ausgegeben werden. Auf der Magnetplatte kann eine Sprachprobe von ca. 140 Sekunden Dauer, auf einem Magnetband von 37 Minuten gespeichert werden. Die Abtastfrequenz beträgt dabei 8 kHz. Durch HARDWARE- und SOFTWARE-Hilfen wird die Benutzung der Sprachein-/ausgabe erleichtert.

5.2 Verwendete Sprachproben

Für die Simulation des PWA-Verfahrens wurden bis zu 128 Sprachquellen benutzt. Jede dieser Quellen wird durch eine Sprachprobe von ca. 17 s Dauer repräsentiert. Die Sprachproben stammen von 8 weiblichen und 13 männlichen Sprechern. Dazu wurden einerseits Nachrichtensendungen und Vorträge vom UKW-Rundfunk übernommen, aber auch Sprachproben aus dem Institut für Angewandte Physik der Universität Frankfurt a.M., die in Deutschland bereits an einigen Universitäten als eine Art Sprachsignalstandard benutzt werden (Sprecher A, ... Sprecher D). In allen Fällen waren oder wurden die Sprachproben durch Bandpaßfilterung auf den für Telefoniequalität relevanten Frequenzbereich zwischen 300 Hz und 3.4 kHz begrenzt.

Bei Messungen wurde (soweit nicht anders angegeben) eine Quelle verwendet, deren Eigenschaften den Mittelwerten über alle Quellen ähnlich war.

Die Sprachproben wurden im Sprachinterface mit 8 kHz abgetastet und in Echtzeit auf ein digitales Magnetband gespeichert. Es handelt sich dabei um durchgehend gesprochenen Text mit einer Aktivität zwischen ca. 75 % und 90 %. Für Untersuchungen an Sprachinterpolationssystemen sind jedoch Sprachproben mit geringerer Aktivität (z.B. 40 %) notwendig. Deshalb wurden die in den Sprachproben vorhandenen Pausen bis zur gewünschten Aktivität künstlich verlängert /5.2-1/ und unter Verwendung eines simulierten Sprachdetektors mit einer Pausenmarkierung versehen.

5.3 Programmierung der Systemsimulation

Für die Sprachsignal-Ein-/Ausgabe in Echtzeit auf Magnetband oder Magnetplatte wurde eine umfangreiche Programmbibliothek erstellt. Dialogorientierte Programme erleichtern dem Benutzer die Bedienung. Durch ein Dateiverwaltungsprogramm wurde das Magnetbandgerät in das plattenorientierte Betriebssystem DOS-M des HP 2100- Rechners einbezogen. Damit ist der Transfer von gespeicherten Sprachsignalen auf Dateiebene möglich, wodurch die Programmierung erheblich vereinfacht wird.

Die eigentliche Programmierung der digitalen Signalverarbeitung erfolgt mit FORTRAN-Programmen. Nur an besonders "exponierten Stellen" werden ASSEMBLER-Unterprogramme zur Verkürzung der Rechenzeit eingesetzt. Dies gilt insbesondere für Datenmanipulationen auf der "Bitebene", die mit Fortran-Programmen nur umständlich möglich sind. Dabei ist wieder ein Kompromiß zwischen Rechenzeit und Übersichtlichkeit des Programms zu schließen. Da besonderer Wert auf eine weitgehend selbstdokumentierende Programmierung gelegt wurde, erhielt die Rechenzeit einen geringeren Stellenwert.

Auf der Magnetplatte können die Signale von 8 Quellen, auf dem Magnetband dagegen von 128 Quellen gespeichert werden. Bei der Simulation mit Scharmittelung über mehr als 8 Quellen muß deshalb immer das Magnetband als Datenquelle verwendet werden. Die Magnetplatte wird dann zur Ablage der Simulationsergebnis-

se verwendet. Aus der seriellen Zugriffsart beim Magnetband ergeben sich bei der Simulation von Sprachinterpolationssystemen Probleme, wenn die 128 Sprachproben einfach hintereinander gespeichert sind. Da zu einem Verarbeitungszeitpunkt die Abtastwerte aller Quellen verfügbar sein müssen, entsteht beim Zusammentragen dieser Werte ein erheblicher Zeitverlust durch die Umspulvorgänge auf dem Magnetband.

Durch einen Umsortiertvorgang der Sprachabtastwerte konnten diese Schwierigkeiten behoben werden. Es wurde von allen Quellen jeweils 64 Abtastwerte zu einem neuen Datenblock (128 * 64 Abtastwerte) zusammengefaßt, der mit einem Lesevorgang in den Kernspeicher geholt werden kann.

Die Rechenzeit für die Bearbeitung einer Quelle beträgt im Mittel 20 Minuten. Dies entspricht bei ca. 17 Sekunden Sprachprobenlänge eine um den Faktor 70 längere Rechenzeit gegenüber einer Echtzeitverarbeitung. Für die Simulation mit 128 Quellen summieren sich die Rechenzeiten auf über 40 Stunden.

6. Ein schneller Signalprozessor für die Echtzeitrealisierung von Sprachverarbeitungssystemen

Bei der Systemsimulation an einem Rechner sollte die Realisierbarkeit einer Idee berücksichtigt werden:

kann ein bestimmter Algorithmus in einem realen System in Echtzeit durchgeführt werden?

wie muß ein Algorithmus vereinfacht werden, daß er in einem bestimmten Zeitabschnitt durchgeführt werden kann und welche Güteverluste sind damit verbunden?

wie groß wird der Aufwand für ein System, das die gegebene Aufgabenstellung in Echtzeit erfüllt?

Von ausschlaggebender Bedeutung für den praktischen Einsatz von Interpolationsverfahren ist der zusätzliche Aufwand an Endeinrichtungen im Vergleich zur Zahl der eingesparten Übertragungskanäle. Darüberhinaus erscheint es unwirtschaftlich, bei einer Simulation Verbesserungen um 1 dB im Störabstand anzustreben, wenn bei einer Hardware-Realisierung durch notwendige Systemvereinfachungen mehrfach größere Verluste auftreten.

Ein Teil der Fragen kann abschließend nur durch eine Realisierung des Systems geklärt werden.

Aus diesem Grund wurde ein Signalprozessor entwickelt und realisiert /6.1-1, ... 6.1-4/, der unter anderem Teile des PWA-Systems in Echtzeit realisieren soll. Bei der Systemplanung konnte auf Erfahrungen mit einem ähnlichen System aufgebaut werden /6.1-5, 6.1-6/.

Im Abschnitt 6.1 wird die Systemarchitektur des Prozessors vorgestellt. Der Abschnitt 6.2 soll die Leistungsfähigkeit aufzeigen. Die Programmiermöglichkeiten werden in 6.3 angesprochen und erste Ergebnisse in 6.4 vorgestellt.

6.1 Architektur des Signalprozessors

Die Forderung an den Signalprozessor nach:

hoher Verarbeitungsgeschwindigkeit und
großer Flexibilität

macht die Entscheidung für bipolare Bit-Slice-Prozessoren leicht. Mit ihnen können Rechnerarchitekturen verwirklicht werden, bei denen eine parallele Verarbeitung mehrerer Operationen in einem Zyklus möglich ist. Dies steht im Gegensatz zu den konventionellen Mikroprozessoren, die immer nur eine Rechenoperation nach der anderen ausführen.

Wichtigster Aspekt des Systems ist die Freiheit bezüglich der Rechnerarchitektur. Dadurch ist es möglich, mehrere von einander unabhängige Funktionseinheiten, gesteuert durch die Mikroprogrammierung, gleichzeitig zu betreiben. Insbesondere wird durch die Auslagerung bestimmter Rechenoperationen aus dem

zentralen Arithmetikprozessor in solche Funktionseinheiten die Verarbeitungsgeschwindigkeit erheblich gesteigert.

Hier einige Beispiele:

Multiplikation durch ein parallel arbeitendes Multiplizierwerk (in 150 ns).

Rotier- und Schiebeoperationen mit einem programmierbaren Shifter (bis zu 16 bit in 150 ns, bis zu 32 bit in 300 ns).

Berechnung von Sinus und Cosinus über eine Tabelle (ROM).

Besondere Bedeutung hat in einem solchen parallel arbeitenden System (Bild 49) der Datentransfer zwischen den einzelnen Funktionseinheiten. Nur durch eine Multibusstruktur kann die Leistungsfähigkeit optimal genutzt werden.

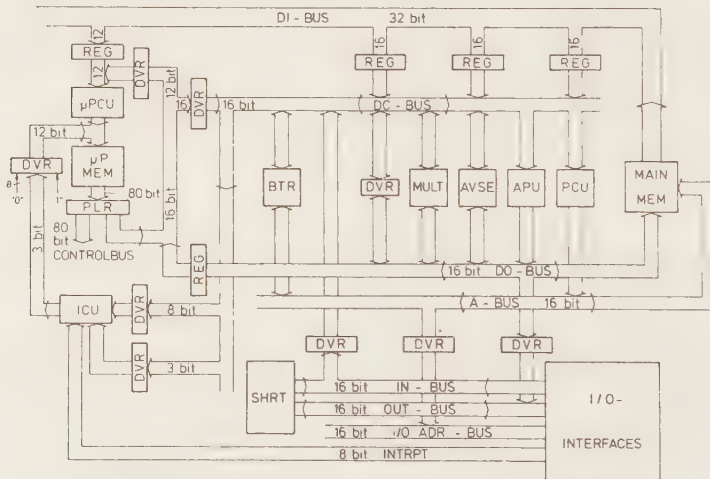


Bild 49: Blockschaltbild des Signalprozessors

REG = Register, μ PCU = Microprogram Control Unit

DVR = BUS-Driver, μ PMEM = Microprogram Memory

PLR = Pipeline Register, ICU = Interrupt Control Unit

BTR = Universal Bidirectional Bus Transceiver

SHRT = Shift Rotate Unit, MULT = 16 Bit Parallel Multiplier

AVSE = Absolute Value Sign and Exponent Evaluation

PCU = Program Control Unit MAIN MEM = Main Memory

APU = Arithmetic Processing Unit

Ebenso wichtig ist die Benutzerfreundlichkeit des Systems. Im vorliegenden Fall erleichtert ein MOS-Mikroprozessor (8085) die Verwaltung des mit RAMs realisierten Mikroprogramm- und Hauptspeichers und ermöglicht Editorfunktionen bei der Mikro- und Makroprogrammierung des Bitscheibenmikroprozessors (Bild 50).

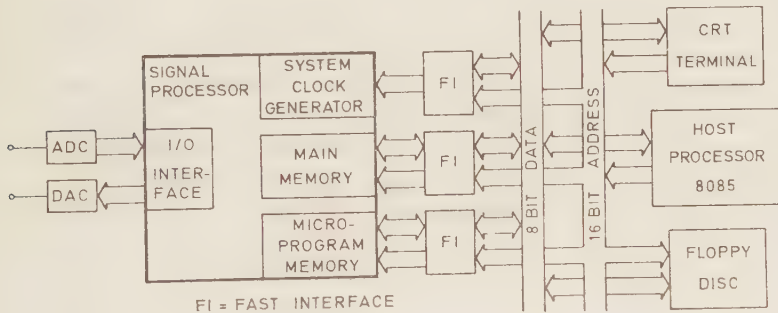


Bild 50: Blockschaltbild des Gesamtsystems

Folgende Besonderheiten des mit Bausteinen der AM 2900 - Familie realisierten Signalprozessors tragen wesentlich zur Leistungsfähigkeit bei:

a) programmierbarer Taktgenerator

Ca. 80 % aller Operationen werden in 150 ns ausgeführt. Für die übrigen kann eine zweite Zykluszeit von 188 ns programmiert werden.

b) Prädiktion von Testbedingungen

Durch Schätzung des voraussichtlichen Ausgangs eines Tests bei bedingten Sprüngen im Mikroprogramm kann gegenüber normalen Pipeline-Strukturen Zykluszeit eingespart werden. Bei falscher Vorhersage wird der aktuelle Zyklus um 112 ns verlängert und das richtige Mikroprogrammwort aus dem Mikroprogrammspeicher

geholt. Die Mikroprogrammierung wird durch dieses Prinzip wesentlich vereinfacht.

c) Adressierung des Hauptspeichers durch zweite APU

Durch die Auslagerung der gesamten Adressrechnung auf einen zweiten Arithmetikprozessor (PCU) wird bei Operationen mit häufigen Hauptspeicherzugriffen oder bei indirekter Adressierung eine kürzere Rechenzeit erreicht.

d) Hauptspeicher auch für MOS-Prozessor während des laufenden Betriebs zugänglich

Der Bitscheibenprozessor greift nicht in jedem Mikroprogrammzyklus auf den Hauptspeicher zu. In diesen Lücken kann der MOS-Prozessor über ein schnelles Interface mit dem Hauptspeicher arbeiten. Damit ist ein Datentransfer zwischen Peripherie und Hauptspeicher ohne Zyklusverlust für den Bitscheibenprozessor verbunden.

6.2 Leistungsfähigkeit des Signalprozessors

Für einen Signalprozessor sind isolierte Rechenzeitangaben (z.B. für die Addition von zwei Registern: 150 ns) nicht aussagekräftig, da dabei Transferzeiten nicht in Erscheinung treten. Sehr viel besser kann die Leistungsfähigkeit des Signalprozessors an konkreten Anwendungsbeispielen aufgezeigt werden. In Bild 51 sind einige, in der digitalen Signalverarbeitung häufig vorkommende, Aufgabenstellungen ausgewählt. Die Abtastfrequenz f gibt den Durchsatz von 16 bit-Worten an.

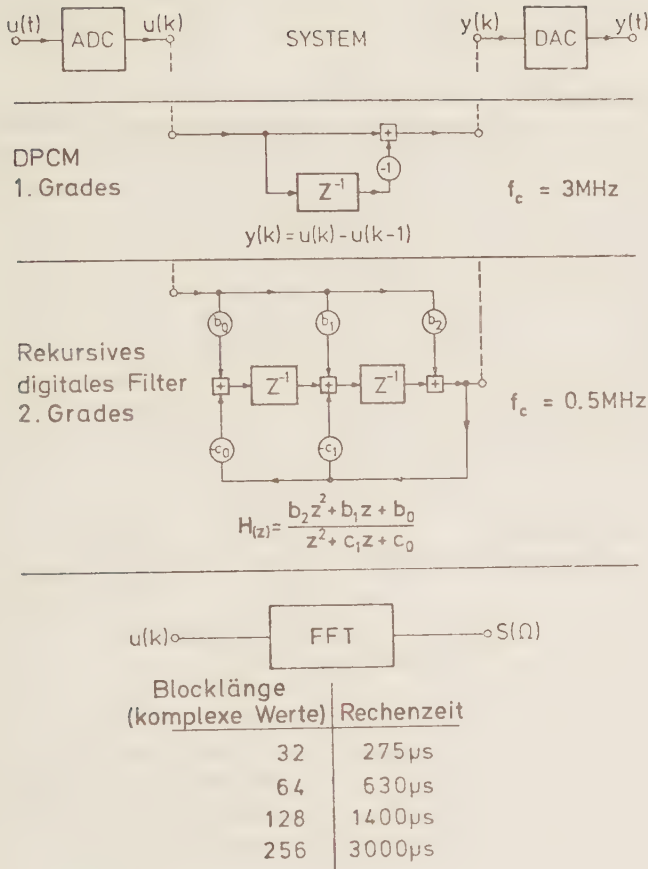


Bild 51: Leistungsfähigkeit des Signalprozessors

6.3 Programmierung des Signalprozessors

Die Programmierung des Signalprozessors kann sowohl auf Mikroprogramm- wie auch auf Makroprogrammebene erfolgen. Die höchste Rechengeschwindigkeit wird erzielt, wenn die Aufgabenstellung durch ein Mikroprogramm gelöst wird. Allerdings ist der Aufwand für die Erstellung eines Mikroprogramms erheblich. Sehr viel eleganter lassen sich die Probleme auf Makroebene bearbeiten. Voraussetzung dafür ist ein entsprechender Standardbefehlssatz. In Verbindung mit einem Makroassembler /6.3-1/ ist auch ohne spezielle Vorkenntnisse über den Signalprozessor eine rasche Programmentwicklung möglich. Neben einem Standardbefehlssatz, der auf den Signalprozessor zugeschnitten ist, wurde ein großer Teil des Befehlssatzes für den Z8000 Mikroprozessor emuliert /6.3 2/. Dadurch ist es möglich, Z8000 Programme auf den Signalprozessor zu übertragen. Am Beispiel eines Cauer-Tiefpasses 4. Grades soll dieser Vorteil durch Vergleich der Rechengeschwindigkeit in Tabelle 5 aufgezeigt werden.

Abtastfrequenz	Prozessor
2.5 kHz	Z8000
12.5 kHz	Signalprozessor mit Z8000-Befehlssatz
200.0 kHz	Signalprozessor (mikroprogrammiert)

Tabelle 5: Vergleich der Rechenzeiten für verschiedene Realisierungen eines Cauer-Tiefpasses 4. Grades

Der Einsatz eines Mikroprogrammassemblys wird zukünftig die Mikroprogrammentwicklung vereinfachen.

6.4 Realisierung von Teilaufgaben im PWA2-System

Es werden hier erste Ergebnisse mitgeteilt, die die Realisierbarkeit von komplexen Sprachübertragungssystemen in den Bereich des Möglichen rückt (wie z B. das PWA-System).

Damit das Signalverarbeitungsproblem in Teilaufgaben zerlegt und an Zwischenergebnissen die Funktion kontrolliert werden kann, ist es zweckmäßig, die Ein- und Ausgangssignale eines Funktionsblockes mit den dazugehörigen Parametern im Hauptspeicher zu halten. Die in Tabelle 6 angegebenen Rechenzeiten sind unter diesem Aspekt zu werten /6.4-1/.

Funktion	Rechenzeit in μ s	Mikroprogramm- speicherbedarf in Byte
lineare Prädiktion mit zwei Koeffizienten	1.5	210
nichtlineare Quantisierungs- kennlinie am Beispiel einer 13-Segmentkennlinie mit um- schaltbarer Wortlänge	2.0	270
Pegelregelung: Sender mit Quantisierung und Empfänger (siehe Bild 30)	35	1070
Sprachdetektor (aus SPEC)	7	470

Tabelle 6: Rechenzeiten für Teilaufgaben beim PWA-Verfahren

7. Zusammenfassung

Es wird ein neues Sprachinterpolationsverfahren mit der Bezeichnung PWA (Sprachinterpolation mit prädiktionsgesteuerter Wortaufteilung) vorgeschlagen.

Eine feste Übertragungskapazität, die aufgrund einer effizienten Codierung und einer Aktivität von 40% im Fernspreverkehr rund 20 kbit/s pro Quelle betragen kann, wird unter die einzelnen Quellen verteilt.

Aufteilungskriterium ist der geschätzte Übertragungsbedarf jeder Quelle, der auf Sender und Empfängerseite aus der Signalvergangenheit berechnet wird. Dadurch entfällt die Signalisierung über die Aufteilung an den Empfänger.

Der zeitlich schwankende Übertragungsbedarf einer Quelle ist abhängig von der möglichen Redundanz- und Irrelevanzreduktion. Als objektives Gütemaß für die Irrelevanzreduktion wird das Toleranzschema nach CCITT verwendet, in dem ein bestimmter Mindeststörabstand für jeden Signalpegel festgelegt ist. Durch lineare Prädiktion wird die Korrelation im Signal reduziert. Der Prädiktor im DPCM-System hat einen festen Koeffizientensatz. In Bereichen geringer Korrelation (z.B. negativer Prädiktionsgewinn) wird auf PCM-Übertragung umgeschaltet.

Das PWA-Verfahren wurde an zwei verschiedenen Beispielen näher untersucht.

Beim PWA1-System, für das kein Sprachdetektor vorgesehen ist und auch eine Übertragung von Daten ausgeklammert wird, benutzt man einerseits den Signalpegel für eine Klassifizierung (zur SNR-Anpassung an das Toleranzschema), andererseits die Leistung von DPCM-Fehlersignalen. Durch diese Klassifizierung können bedingte Häufigkeitsverteilungen von PCM- und DPCM-Werten ermittelt werden. Ein einfacher aber leistungsfähiger Algorithmus liefert in einem iterativen Verfahren näherungsweise optimale Kennlinien. Bei einer Sprachaktivität von 40 % beträgt die mittlere Bitrate bis zu 3.2 bit pro Abtastwert. Der Störabstand liegt bei 35 dB. Auch in den Pausen muß mindestens 1 bit/Abtastwert übertragen werden. Diese relativ hohe Daten-

rate in den Pausen hat zur Entwicklung des PWA2-Systems geführt. Durch segmentweisen Pausenausschluß (z.B. jeweils 128 Abtastwerte) mit einem Signalisierbit pro Segment sinkt die Datenrate in den Pausen auf $1/128$ bit pro Abtastwert. Mit einem weiteren Steuerbit pro Segment kann auch eine ungestörte Datenübertragung erzwungen werden. Damit ist als Signalisierungsinformation $1/64$ bit pro Abtastwert aufzuwenden.

Bei einer Sprachaktivität von 40% kann bei Scharmittelung über 128 Quellen bereits mit einer Datenrate von 22 kbit/s das CCITT-Toleranzschema erfüllt werden.

Aber auch bei Datenraten von 20 kbit/s sind keine hörbaren Störungen feststellbar. Erst bei 16 kbit/s und bei niedrigen Quellenzahlen (z.B. 16) sind geringe Störungen wahrnehmbar. Es handelt sich dabei um eine rauschähnliche Störung. Mit der Segmentierung sind weitere Vorteile verbunden, die bei Maßnahmen gegen Übertragungsfehler in Erscheinung treten.

Die Untersuchungen wurden mittels Computersimulation am Institutsrechner durchgeführt.

Voraussetzung dafür ist ein Ein-/Ausgabegerät für Analogsignale am Rechner. Das Sprachinterface, für diesen Zweck entwickelt, kann Analogsignale mit 12 bit linear quantisieren und mit einer Abtastfrequenz von 35 kHz in Echtzeit an den Rechner übergeben. Von dort aus ist eine Speicherung auf Magnetplatte oder Magnetband in Echtzeit möglich. Die notwendige Software, auch für die Echtzeitsignalausgabe von den gleichen Speichermedien, wurde erstellt.

Schon in einem früheren Stadium wurden Realisierungsüberlegungen zu den entwickelten Sprachinterpolationssystemen angestellt. Um zuverlässige Abschätzungen zu ermöglichen, wurde deshalb ein leistungsfähiger Signalprozessor erstellt.

Dieser Signalprozessor benötigt für die Berechnung der Prädiktion 2. Grades $1\mu\text{s}$, für eine 13-Segment-Codierung und anschließender Decodierung $2\mu\text{s}$.

Zum Vergleich sei noch die Rechenzeit eines Cauer-Tiefpasses 4. Grades mit $5\mu\text{s}$ und die Transformation von 256 komplexen Werten mit der FFT in 3 ms angegeben.

Im Rahmen dieser Arbeit konnten einige Probleme zwar erkannt, aber im Interesse eines endlichen Umfangs nicht, oder nur unvollständig bearbeitet werden.

Dazu zwei Beispiele, die zu weiterführenden Arbeiten anregen sollen.

An erster Stelle wäre der Einfluß von Störungen auf dem Übertragungskanal zu nennen. Dazu sind weitere Untersuchungen erforderlich.

Ebenfalls sehr interessant ist ein Modell, mit dem die Beeinflussung der von einer Quelle angeforderten Wortlänge beschrieben werden kann. Ein Ziel wäre, diese durch Sprachinterpolation mit endlicher Quellenzahl bedingte Wortlängenreduktion oder -vergrößerung, in Form eines geeigneten Zufallsprozesses zu realisieren. Dann wäre man in der Lage, in einem System mit nur einer Quelle, die Effekte eines Systems mit n Quellen zu simulieren.

Herrn Professor Dr.-Ing. D. Seitzer danke ich für die interessante Aufgabenstellung, für zahlreiche Anregungen und für die Unterstützung bei der Durchführung dieser Arbeit.

Herrn Professor Dr.-Ing. W. Schüller gebührt mein Dank für die Übernahme der Mitberichterstattung.

Den Herren Dr.-Ing. G.G. Langenbacher, Dr.-Ing. F. Oehme und Dipl.-Ing. H. Rauch danke ich für zahlreiche fruchtbare Diskussionen.

Danken möchte ich auch den Herren Dipl.-Ing. D. Bauer und Dipl.-Ing. H. Schramm, die wesentlich zum Gelingen des Signalprozessorprojekts beigetragen haben und last not least meiner Frau für die mühevollen Kleinarbeit bei der Erstellung der Zeichnungen und bei der Reinschrift des Manuskripts.

Literaturverzeichnis

- /2.1-1/ Fano, R.M.
Informationsübertragung
Eine statistische Theorie der Nachrichtenübertragung
Oldenburg Verlag München-Wien 1966

- /2.1-2/ Musmann, H.G.:
Theoretische Grundlagen der Quellencodierung
Technische Probleme zukünftiger Breitband-
Nachrichtennetze. Konferenz im Fernmeldetechnischen
Zentralamt, Darmstadt 1973
Tagungsband S. 174-186

- /2.1-3/ Shouten, J.F.:
Nachricht und Signal
Nachrichtentechnische Fachberichte, Bd. 6, 1975
S. II.1-II.2

- /2.1-4/ Makhoul, J.:
Linear Prediction: A Tutorial Review
Proceedings of the IEEE, Vol. 63, No. 4, April 1975

- /2.1-5/ Gallager, R.G.:
Information Theory and Reliable Communication
John Wiley & Sons, Inc. 1968 New York

- /2.1-6/ Huffman, D.A :
A Method for the Construction of Minimum Redundancy
Codes
Proc. IRE 40, 1952, pp. 1098-1101

- /2.1-7/ Berger, T.:
Rate Distortion Theory
Prentice-Hall, 1971

- /2.1-8/ Goodman, D.J.; Goodman, J.S.; Chen, M.:
Intelligibility and Ratings of Digitally Coded
Speech
IEEE Transactions on Acoustics, Speech and
Signal Processing Vol. ASSP-26, No. 5, October 1978
- /2.2-1/ Cattermole, K.W.:
Principles of Pulse Code Modulation
London Iliffe Books Ltd., 1969
- /2.2-2/ Schüßler, W.:
Nachrichtenübertragung
Vorlesung an der Universität Erlangen-Nürnberg
- /2.2-3/ Croisier, A.; Jacquart, C.:
A Single Channel P.C.M. Coder
Proc. Internat. Zurich Seminar 1972 S. F2(1)-F2(4)
- /2.2-4/ Opitz, L.:
Vielfach-TDM-Fernsprecheinrichtung mit
30/32-Kanal-PCM-System
Standard Elektrik Lorenz AG
- /2.2-5/ Noll, P.:
Über einige Eigenschaften von Differenz-PCM-Systemen
A E Ü, Band 30, 1976, H. 3, S. 125 130
- /2.2-6/ Noll, P.:
Sprachübertragung mit adaptiven
Differenz-PCM-Verfahren
Techn. Bericht Nr. 164, Heinrich-Hertz-Institut,
Berlin 1973

- /2.2-7/ Jayant, N.S.:
Digital Coding of Speech Waveforms:
PCM, DPCM, and DM Quantizers
Proc. of the IEEE Vol. 62, No. 5, May 1974

- /2.2-8/ Dietrich, M.:
Coding of Speech Signals Using a Switched Quantizer
Proc. 1974, Intern. Zürich Seminar on Digital
Communications, Zürich, March 1974

- /2.2-9/ Langenbucher, G.G.:
Bestimmung von optimalen Prädiktionskoeffizienten
an Sprachproben für die Simulation von Sprachüber-
tragungsverfahren (persönliche Mitteilung)

- /2.2-10/ Musmann, H.G.: Über lineare Transformation zur
Redundanzreduktion
Nachrichtentechnische Fachberichte Codierung
Band 40, 1971, VDE-Verlag GmbH

- /2.2-11/ Zelinski, R.; Noll, P.:
Adaptive Blockquantisierung von Sprachsignalen
Techn. Bericht Nr. 181, Heinrich-Hertz-Institut,
Berlin 1975

- /2.2-12/ Andrews, C A.; Davies, J.M.; Schwarz, G.R.:
Adaptive Data Compression
Proc. of the IEEE 55, 1967 H. 3, S. 267-277

- /2.2-13 Zelinski, R.; Noll, P.:
Adaptive Transform Coding of Speech Signals
IEEE Transactions on Acoustics, Speech and
Signal Proc., Vol. ASSP-25, No. 4, August 1977

- /2.2-14/ Crochiere, R.E.; Webber, S.A.; Flanagan J.L.:
Digital Coding of Speech in Sub-bands
Bell System Technical Journal Vol. 55, No.8
October 1976
- /2.2-15/ Endres, W.:
Methoden der Sprachanalyse und Sprachsynthese
und ihre Bedeutung für die Nachrichtentechnik
Der Fernmelde-Ingenieur 11. Jahrgang, Heft 11,
November 1957
- /2.2-16/ Flanagan, D.J.; Goodman J.S.; Chen, M.;
Crochiere, R.E.; Jayant, N.S.; Tribolet, J.N.:
Speech Coding
IEEE Trans. Communications Vol. COM-27, No.4, 1979
- /2.3-1/ Langenbucher, G.G.:
Messungen an Sprachproben mit definierter Aktivität
Persönliche Mitteilung
- /2.3-2/ Bullington, K.; Fraser, J.M.:
Engineering Aspects of TASI
Bell Syst. Techn. J., March 1959
- /2.3-3/ Leopold, G.R.:
TASI-B: A System for Restauration and
Expansion of Overseas Circuits
Bell Laboratories Record 10, 1970, S. 299 306
- /2.3-4/ Amano, K.; Ota, C.; Denskin, K.:
Digital TASI System in PCM-Transmission
Intern. Conf. on Communications 1969
Conf. Rec. S. 34.23-34.29

- /2.3-5/ Penicaud, E.; Salle, Y.; Guillarme, P.:
Concentrateur numerique de circuits telephoniques
a grande capacite pour station AMRT (CELTIC)
L'onde electrique 53 (1973) 1, S.25-29

- /2.3-6/ Sciulli, J.A.; Campanella, S.J.:
A Speech Predictive Encoding Communication
System for Multichannel Telephony
IEEE Trans. on Communications Vol. COM-21,
No. 7, July 1973, S. 827-835

- /2.3-7/ Weitowitz, R.:
Ein digitales Sprachinterpolationsverfahren
mit momentaner Prioritätszuteilung
Dissertation am Fachbereich Ingenieurwissenschaften
der Universität Erlangen-Nürnberg 1977

- /2.3-8/ Seitzer, D.; Closs, F.; Stucki, P.:
An Experimental Approach to Video Bandwidth
Compression by Multiplexing
IEEE Trans. on Communications COM-17, 1969,
5, S. 564-568

- /2.3-9/ Seitzer, D.:
Instantaneous Priority Multiplexing of
Gray Pictures
IEEE Trans. on Communications COM-18,
1970, 5, S. 690-692

- /2.3-10/ Oehme, F.:
Statistische Untersuchungen zur Bildcodierung
nach dem Überlagerungsprinzip mit momentaner
Prioritätszuteilung
Dissertation am Fachbereich Ingenieurwissenschaften
der Universität Erlangen-Nürnberg 1975

- /2.3-11/ Tietze, U.:
Untersuchung eines datenreduzierenden Multiplex-
Verfahrens für die digitale Bildübertragung
Dissertation am Fachbereich Ingenieurwissenschaften
der Universität Erlangen-Nürnberg 1975
- /2.3-12/ Hess, H.:
Untersuchung datenreduzierender Multiplexverfahren
zur digitalen Übertragung von Bildern mit zwei
Helligkeitsstufen
Dissertation am Fachbereich Ingenieurwissenschaften
der Universität Erlangen-Nürnberg 1977
- /2.3-13/ Acs, E.; Hutter, O.:
Eine 30-Kanal-Multiplexeinrichtung nach dem
lagemodulierten Adressen-Kode-System
Nachrichtentechnik 17, 1967, 2, S. 55-58
- /2.3-14/ Poretti, I.; Monty, G.; Bagnoli, A.:
Speech Interpolation Systems and their
Application in TDM/TDMA Systems
Telecommunications Numeriques par
Satellite, S. 348-357 Traitement du Signal
- /2.3-15/ Molinari, A.M.; Vagliani, F.C.:
Bit Rate per Channel Halving in PCM
Multiplexes by Speech Interpolation and
Adaptive Quantization
Proc. Internat. Zurich Seminar on Digital
Communications 1976 S.A7.1 A7.8

- /2.3-16/ Dorgelo, A.J.G.; Van der Veer, H.:
Variable Length Coding Increasing
Traffic Capacity in PCM Transmission
Systems
Proc Internat. Zurich Seminar on Integrated
Systems for Speech, Video and Data
Communications 1972, S.C6.1-C6.6
- /2.3-17/ Gatefield, A.G.; Suyderhoud, H.G.;
Wolejsza, C.J.:
System Design for the Digitally Implemented
Communication Experiment (DICE)
Internat. Conf. on Communications, Chicago
1977, Conf. Rec. Vol.I, S.1-5
- /2.3-18/ Gerhäuser, H.; Seitzer, D.; Weitowitz, R.:
Ein Zeitmultiplexverfahren mit prädiktions-
gesteuerter Wortaufteilung
Nachrichtentechnische Zeitschrift 28 (1975) 1,
S. 25-27
- /2.3-19/ Langenbucher, G.G.:
Ein Verfahren zur Verringerung der Datenrate
bei der Sprachübertragung durch blockweise
Prioritätszuteilung (BPZ)
Nachrichtentechn. Zeitschrift 29 (1976) 11,
S. 809-810
- /2.3-20/ Langenbucher, G.G.:
Ein digitales Sprachinterpolationsverfahren
mit blockweiser Prioritätszuteilung
Dissertation am Fachbereich Ingenieur-
wissenschaften der Universität Erlangen-
Nürnberg (1978)

- /2.3-21/ Mc Pherson, T.R.J.; O'Neal, J.B.Jr.;
Stroh, R.W.:
PCM Speech Compression via ADPCM/TASI
Internat. Conf. on Acoustics, Speech and
Signal Processing, 1977
- /2.3-22/ Seitzer, D.; Gerhäuser, H.; Langenbucher, G.G.;
Woitowitz, R.:
Fortschritte bei der digitalen Sprachinterpolation
Nachrichtentechnische Zeitschrift 30 (1977) 11,
S.849-854
- /2.3-23/ Seitzer, D.; Gerhäuser, H.; Langenbucher, G.G.:
High Quality Digital Speech Interpolation Methods
Internat. Conf. on Communications Toronto 1978,
Conf. Rec. Vol.III, S.50.2.1-50.2.5
- /2.3-24/ Elder, J.M.; O'Neill, J.F.:
A Speech Interpolation System for Private Networks
Proc. Nat. Telecom. Conf. 1978
Birmingham, Alabama
- /3.3-1/ CCITT Orange Book Vol. III-2 Rec. G.711
- /4.1-1/ Max, J.:
Quantizing for minimum distortion
IRE Trans. on Information Theory
IT-6 (March 1960) pp. 7-12
- /4.1-2/ Panther, P.F.; Dite, W.:
Quantization Distortion in Pulse-Count
Modulation with Nonuniform Spacing of Levels
Proc. of IRE 39 (Jan. 1951) pp. 44-48

- /4.1-3/ Cohen, F.A.:
Quantisierung und Codierung von Markoff-
Quellen angewandt auf Sprachsignale
Dissertation an der TU Braunschweig 1975
- /4.2-1/ Noll, P.:
Adaptive Quantizing in Speech Coding Systems
Internat. Zurich Seminar on Digital
Communication 1974
Conf. Rec. PP. B3 (1)...(6)
- /4.2-2/ Wolf, D.:
Analytische Beschreibung von Sprachsignalen
A E Ü Band 31 (1977) Heft9 S.392...398
- /4.2-3/ Langenbucher, G.G.:
Anwendung der Sprachdetektion auf gespeicherte
Sprachproben
Bericht über ein von der Deutschen Forschungs-
gemeinschaft gefördertes Forschungsvorhaben 1979
- /5.1-1/ Gerhäuser, H.:
Aufbau eines Interface zur Eingabe von
Sprachsignalen in den Computer HP2100A.
Durchführung von statistischen Unter-
suchungen an Sprachsignalen mit Hilfe
des Computers
Diplomarbeit am Lehrstuhl für Technische
Elektronik der Universität Erlangen-
Nürnberg 1973
- /5.2-1/ Langenbucher, G.G.:
Erstellung von Sprachproben mit variabler Belegung
Bericht über ein von der Deutschen Forschungs-
gemeinschaft gefördertes Forschungsvorhaben 1975

- /6.1-1/ Mühlberger, K.:
Realisierung eines schnellen Bitscheiben-
Mikroprozessors für die Echtzeitsignalverarbeitung
Diplomarbeit am Lehrstuhl für Technische
Elektronik der Universität Erlangen-Nürnberg 1978
- /6.1-2/ Reichmeyer, H.:
Einsatz eines MOS-Mikroprozessors (8085) zur
Verwaltung eines Bitscheibenprozessors
Diplomarbeit am Lehrstuhl für Technische
Elektronik der Universität Erlangen-Nürnberg 1978
- /6.1-3/ Bauer, D.:
Realisieren einer schnellen externen Multiplizier-
Akkumuliereinheit für einen Bitscheiben-
Mikroprozessor-System
Diplomarbeit am Lehrstuhl für Technische Elektronik
der Universität Erlangen-Nürnberg 1979
- /6.1-4/ Gerhäuser, H.:
A Fast Bit Slice Computer for Real-Time
Signal Processing
EUROCON 80 Stuttgart, Conf. Rec. S.312-314
- /6.1-5/ Schloß, J.; Müller, H.:
Ein mikroprogrammierbares Entwicklungssystem
für bipolare Mikroprozessorelemente
Elektronik 1978 H.8 S40-47
- /6.1-6/ Kolb, H.J.
A Signal Processor Using Bitslice Elements
for the Audio Frequency Range
Signal Processing, October 1980

- /6.3-1/ Schramm, H.:
Ein Makroassembler für das Z8000 Entwicklungssystem
AMSYS8/8 der Firma Advanced Micro Devices
Persönliche Mitteilung
- /6.3-2/ Pirkl, G.; Schramm, H.:
Emulation von Z8000-Befehlen am Bitscheiben-
Signalprozessor
Persönliche Mitteilung
- /6.4-1/ Bauer, D.:
Rechenzeiten für Teilaufgaben
aus dem PWA-Verfahren
Persönliche Mitteilung

9. Anhang

Es soll der Zusammenhang zwischen Segmentleistungsstreuung S_L und der Segmentlänge L aufgezeigt werden.

Zunächst wird der Begriff "Segmentleistungsstreuung" präzisiert. Es soll darunter die Standardabweichung (positive Wurzel aus der Varianz) der Segmentleistungen verstanden werden. S_L hat die Dimension einer Leistung.

Voraussetzungen:

Stochastisches Signal $u(k)$, mittelwertfrei, Abtastwerte statistisch unabhängig, Gaußsche Amplitudendichteverteilung (AVD).

Die Gleichung für die Segmentleistung lautet:

$$y_L(k) = \frac{1}{L} \sum_{i=0}^{L-1} u^2(k+i) \quad (1)$$

$$y_L(k) = \frac{1}{L} \left[u^2(k) + u^2(k+1) + \dots + u^2(k+L-1) \right]$$

Da die Abtastwerte $u(k)$ statistisch unabhängig sind, kann der Satz von der Summe der Varianzen angewendet werden:

$$\sigma_z^2 = \sum_{i=1}^L \sigma_i^2 \quad (2)$$

(σ_i^2 = Varianzen der Zufallsvariablen x_i)

σ_z^2 = Varianz des Summenprozesses $Z = X_1 + X_2 + \dots + X_L$)

Sind alle $\sigma_i^2 = \sigma_0^2$, dann gilt:

$$\sigma_z = \sigma_0 \sqrt{L} \quad (3)$$

Umbenennung:

$$x(k) = u^2(k) \quad (4)$$

$$\bar{x} = E \{ x(k) \} = E \{ u^2(k) \} = \sigma_u^2 \quad (5)$$

($u(k)$ ist mittelwertfrei; σ_u^2 = Signalleistung S_0)

(4) in (1) eingesetzt:

$$y_L(k) = \frac{1}{L} \sum_{i=0}^{L-1} x(k+i) \quad (6)$$

Die Standardabweichung von $x(k)$ sei S_x .

Dann kann mit (3) und (6) die Standardabweichung S_L von y_L beschrieben werden:

$$S_L = \frac{\sqrt{L}}{L} S_x$$

$$S_L = \frac{1}{\sqrt{L}} S_x \quad (7)$$

Die Streuung der Segmentleistungen verhält sich umgekehrt proportional zur Wurzel aus der Segmentlänge L .

Für $L = 1$ vereinfacht sich (6):

$$y_{L=1}(k) = x(k) = u^2(k)$$

(7) kann geschrieben werden:

$$S_{L=1} = S_x$$

$$S_x^2 = E \left\{ [x(k) - \bar{x}]^2 \right\}$$

$$= E \left\{ x^2(k) \right\} - 2 \underbrace{\bar{x} E \left\{ x(k) \right\}}_{\bar{x}} + \bar{x}^2$$

$$S_x^2 = E \left\{ x^2(k) \right\} - \bar{x}^2 \quad (8)$$

Normiert man auf $\bar{x}^2$ dann wird:

$$\frac{S_x^2}{\bar{x}^2} = \frac{E \left\{ x^2(k) \right\}}{\bar{x}^2} - 1 \quad (9)$$

Der Variationskoeffizient v ist definiert als:

$$v = \frac{S_x}{\bar{x}} \quad (10)$$

mit (9):

$$v = \sqrt{\frac{E \{x^2(k)\}}{\bar{x}^2} - 1}$$

Setzt man (4) und (5) ein:

$$v = \sqrt{E \left\{ \frac{u^4(k)}{\sigma_u^4} \right\} - 1} \quad (11)$$

In (11) steht das 4. Moment der Zufallsvariablen u in Standardform.

$$E \left\{ \left(\frac{u}{\sigma} \right)^4 \right\} = \int_{-\infty}^{\infty} \left(\frac{u}{\sigma} \right)^4 p(u) du \quad (12)$$

$p(u)$ war als Gaußsche AVD vereinbart, deshalb:

$$\int_{-\infty}^{\infty} \left(\frac{u}{\sigma} \right)^4 p(u) du = 3 \quad (13)$$

Eingesetzt in (11):

$$v = \sqrt{3 - 1}$$

$$v = \sqrt{2} \quad (\text{für } L = 1)$$

mit (7) kann man dann allgemein schreiben:

$$v = \frac{\sqrt{2}}{\sqrt{L}}$$

Ergebnis:

Der Variationskoeffizient v von Segmentleistungsstreuungen nimmt abhängig von der Segmentlänge L mit $1/\sqrt{L}$ ab.
Bei $L = 1$ wird $v = \sqrt{2}$.

10. Verwendete Abkürzungen und Formelzeichen

a	Aktivität einer Quelle
a_i	Prädiktionskoeffizient i
AVD	Amplitudenverteilungsdichte
ATC	Adaptive Transformationscodierung
ATW	Abtastwert
ATIC	Allocazione nel Tempo con Interpolatione di Campionatura
AWL	angeforderte Wortlänge
APD	angeforderte PCN/DPCM-Umschaltung
AKLN	angeforderte Kennliniennummer
B	Systembelastung
B_n	auf n Quellen normierte Systembelastung
BPZ	Blockweise Prioritätszuteilung
$E(k)$	Einhüllende eines Signals
f_A	Abtastfrequenz
f_g	Grenzfrequenz
f_m	mittlere Frequenz
f_z	Nulldurchgangsrate
$F(x)$	Verteilungsfunktion, Summenhäufigkeit
FFT	Fast Fourier Transform
CCITT	Comité Consultatif International Téléphonique et Télégraphique
C_v	Normierungskonstante bei der Kennlinienzuordnung
DSI	Digitale Speech Interpolation
DPCM	Differenzpulsmodulation
DM	Deltamodulation
$d(k)$	DPCM-Fehlersignal
$d_i(k)$	quantisiertes DPCM-Fehlersignal
$D^*(k)$	Schätzwert für Kurzzeiteffektivwert von DPCM- Werten
$\Delta d(k)$	Quantisierungsfehler
DST	Diskrete Slant-Transformation
DFT	Diskrete Fourier Transformation
DCT	Diskrete Cosinus Transformation
DPCM-Wert	DPCM-Fehlersignal $d(k)$

Gp	Prädiktionsgewinn	DPCM-Werte
G11		kleine
G12	Signalgruppen in der Pegelklasse KS1	mittlere
G13		große
G21		
G22	Signalgruppen in der Pegelklasse KS2	
G23		
G31		
G32	Signalgruppen in der Pegelklasse KS3	
G33		
H ₀	Nachricht, Information	
H	Informationsgehalt, Entropie	
I	Quantisierungsintervallgrenze	
K	ganzzahliger Repräsentant	
k	Nr. des Abtastzeitpunkts	
KLN	Kennlinien-Nr.	
KS1		kleine Signalpegel
KS2	Signalpegelklassen	mittlere Signalpegel
KS3		große Signalpegel
KLT	Karhunen Loeve-Transformation	
LDM	Lineare Deltamodulation	
LSB	Least Significant Bit	
L	Segmentlänge (Zahl der Abtastwerte in einem Segment)	
ld	Segmentlänge für Berechnung von $D^*(k)$	
ls	Segmentlänge für Berechnung von $S^*(k)$	
MSB	Most Significant Bit	
MPZ	Momentane Prioritätszuteilung	
N, N _X	Zahl der Quantisierungsstufen	
N1, N2		
n	Zahl der Quellen in einem Sprachinterpolations- system	
PCM	Pulscodemodulation	
PCM-Werte	Signalabtastwerte $u(k)$	
PWA	Prädiktionsgesteuerte Wortaufteilung	
13-Seg.PCM	PCM mit nichtlinearer Kompondierung unter Verwendung einer 13-Segmentkennlinie	
p	Wahrscheinlichkeitsdichte	
P	Wahrscheinlichkeit	
PCM-RR	PCM with Rate and Redundancy Reduction	
Q ₀	Quantisierungsfehlerleistung	

Q	Fehlerleistung im μ -ten Quantisierungsintervall
$q(k)$	pegelgeregelte Signalabtastwerte
r_{xy}	Korrelationskoeffizient
$r(t)$	Rauschsignal
r	Repräsentant
R	Redundanz
R_D	Datenrate
$R_D(d)$	Rate Distortion Funktion
S	Signalpegel
$s(t)$	Sprachsignal
$S^*(k)$	Schätzwert für Kurzzeiteffektivwert von PCM-Werten
SIDEL	Sprachinterpolation mit Deltamodulation
SNR	Signal to Noise Ratio, Störabstand
Δ SNR	SNR-Abstand zum Toleranzschema
SPEC	Speech predictive Encoding
S_L	Standardabweichung der Segmentleistungen
σ_0	Signaleffektivwert
$\sigma(k)$	Kurzzeiteffektivwert
SKS1	Klassengrenze zwischen KS1 und KS2
SKS3	Klassengrenze zwischen KS2 und KS3
SG11	Gruppengrenze zwischen G11 und G12
SG13	G12 und G13
SG21	Gruppengrenze zwischen G21 und G22
SG23	G22 und G23
SG31	Gruppengrenze zwischen G31 und G32
SG33	G32 und G33
σ_d	Varianz des Prädiktionsfehlers
ZWL	zugeteilte Wortlänge
ZPD	zugeteilte PCM/DPCM-Umschaltung
ZKLN	zugeteilte Kennlinien-Nr.

Lebenslauf

- 24.5.1946 Geboren in München als Sohn des Diakons Ludwig Gerhäuser und seiner Ehefrau Margarete Gerhäuser, geb. Mayer.
- 1952-1960 Besuch der Volksschule in Rummelsberg und Fürth
- 1960-1964 Berufsausbildung zum Elektromechaniker
Abschluß: Facharbeiterbrief
- 1962-1965 Berufsaufbauschule
Abschluß: Mittlere Reife
- 1965-1968 Besuch der Ingenieurschule Ohm-Polytechnikum-Nürnberg
Abschluß: Ing. (grad.)
- 1968 Fachgebundene Hochschulreife
- 1968-1973 Elektrotechnikstudium an der Universität Erlangen-Nürnberg
Abschluß: Dipl.-Ing.
- seit 1973 Wissenschaftlicher Assistent am Lehrstuhl für Technische Elektronik der Universität Erlangen-Nürnberg

Erlangen, April 1980

Heinz Gerhäuser

